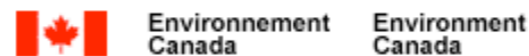




Atelier MITACS sur la *Diffusion multi-format
d'informations environnementales*

Analyse et consolidation d'information de la base de la base Météocode

Pascal Vincent & Sébastien Langevin





Laboratoire d'Informatique des Systèmes Adaptatifs

Professeurs



Yoshua BENGIO



Douglas ECK



Pascal VINCENT



- 3 professeurs
- 1 chercheur sénior et 2 post-docs
- 12 étudiants au Doctorat
- 10 étudiants à la Maîtrise

Notre expertise:

- ★ Apprentissage Statistique (Machine Learning)
- ★ Data-mining (analyse de grands ensembles de données)
- ★ Réseaux de neurones artificiels

Yoshua Bengio:

Chaire de Recherche du Canada sur les Algorithmes d'Apprentissage Statistique

Chaire de recherche industrielle CRSNG-CGI en fouille de données en haute dimension pour les finances électroniques

L'apprentissage statistique

en quelques mots

- Apprendre une *fonction* effectuant des **prédictions**
- En se basant sur un **grand ensemble d'exemples** (base de données)
- Capable de **généraliser** à de nouveaux **cas** (pas vus dans les données)

L'apprentissage statistique (supervisé)

L'apprentissage statistique (supervisé)

Ensemble de données d'entraînement (*training set*)

L'apprentissage statistique (supervisé)

Ensemble de données d'entraînement (*training set*)

1	entrée, observation, <i>input</i> , x_1	cible, <i>target</i> , sortie désirée, y_1
2	entrée, observation, <i>input</i> , x_2	cible, <i>target</i> , sortie désirée, y_2
3	entrée, observation, <i>input</i> , x_3	cible, <i>target</i> , sortie désirée, y_3
⋮	<i>etc...</i>	⋮
n	entrée, observation, <i>input</i> , x_n	cible, <i>target</i> , sortie désirée, y_n

L'apprentissage statistique (supervisé)

Ensemble de données d'entraînement (*training set*)

1	entrée, observation, <i>input</i> , x_1	cible, <i>target</i> , sortie désirée, y_1
2	entrée, observation, <i>input</i> , x_2	cible, <i>target</i> , sortie désirée, y_2
3	entrée, observation, <i>input</i> , x_3	cible, <i>target</i> , sortie désirée, y_3
⋮	<i>etc...</i>	⋮
n	entrée, observation, <i>input</i> , x_n	cible, <i>target</i> , sortie désirée, y_n

taille de
l'ensemble,
nombre
d'exemples,
d'échantillons.

L'apprentissage statistique (supervisé)

Ensemble de données d'entraînement (*training set*)

1	entrée, observation, <i>input</i> , x_1	cible, <i>target</i> , sortie désirée, y_1
2	entrée, observation, <i>input</i> , x_2	cible, <i>target</i> , sortie désirée, y_2
3	entrée, observation, <i>input</i> , x_3	cible, <i>target</i> , sortie désirée, y_3
⋮	<i>etc...</i>	⋮
n	entrée, observation, <i>input</i> , x_n	cible, <i>target</i> , sortie désirée, y_n

point de test

taille de
l'ensemble,
nombre
d'exemples,
d'échantillons.

L'apprentissage statistique (supervisé)

Ensemble de données d'entraînement (*training set*)

1	entrée, observation, <i>input</i> , x_1	cible, <i>target</i> , sortie désirée, y_1
2	entrée, observation, <i>input</i> , x_2	cible, <i>target</i> , sortie désirée, y_2
3	entrée, observation, <i>input</i> , x_3	cible, <i>target</i> , sortie désirée, y_3
⋮	<i>etc...</i>	⋮
n	entrée, observation, <i>input</i> , x_n	cible, <i>target</i> , sortie désirée, y_n

point de test



? ? ?

taille de l'ensemble, nombre d'exemples, d'échantillons.

L'apprentissage statistique (supervisé)

Ensemble de données d'entraînement (*training set*)

1	entrée, observation, <i>input</i> , x_1	cible, <i>target</i> , sortie désirée, y_1
2	entrée, observation, <i>input</i> , x_2	cible, <i>target</i> , sortie désirée, y_2
3	entrée, observation, <i>input</i> , x_3	cible, <i>target</i> , sortie désirée, y_3
⋮	<i>etc...</i>	⋮
n	entrée, observation, <i>input</i> , x_n	cible, <i>target</i> , sortie désirée, y_n

point de test



???

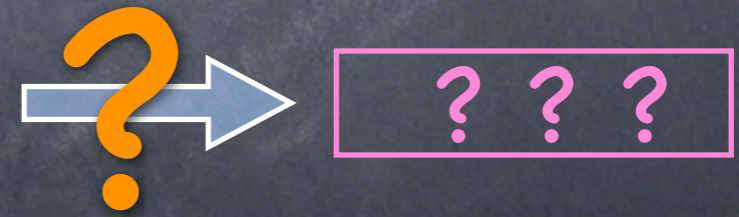
taille de l'ensemble, nombre d'exemples, d'échantillons.

L'apprentissage statistique (supervisé)

Ensemble de données d'entraînement (*training set*)

1	entrée, observation, <i>input</i> , x_1	cible, <i>target</i> , sortie désirée, y_1
2	entrée, observation, <i>input</i> , x_2	cible, <i>target</i> , sortie désirée, y_2
3	entrée, observation, <i>input</i> , x_3	cible, <i>target</i> , sortie désirée, y_3
⋮	<i>etc...</i>	⋮
n	entrée, observation, <i>input</i> , x_n	cible, <i>target</i> , sortie désirée, y_n

point de test



On cherche un algorithme qui produit une sortie (*output*) qui est une **bonne prédiction** de la cible. Cet algorithme trouve une bonne fonction $x \rightarrow y$

taille de l'ensemble, nombre d'exemples, d'échantillons.

L'apprentissage statistique (supervisé)

Ensemble de données d'entraînement (*training set*)

Une **entrée** est généralement représentée par un vecteur de dimension **d**.

1	entrée, observation, <i>input</i> , x_1	cible, <i>target</i> , sortie désirée, y_1
2	entrée, observation, <i>input</i> , x_2	cible, <i>target</i> , sortie désirée, y_2
3	entrée, observation, <i>input</i> , x_3	cible, <i>target</i> , sortie désirée, y_3
⋮	<i>etc...</i>	⋮
n	entrée, observation, <i>input</i> , x_n	cible, <i>target</i> , sortie désirée, y_n

point de test



???

taille de l'ensemble, nombre d'exemples, d'échantillons.

On cherche un algorithme qui produit une sortie (*output*) qui est une **bonne prédiction** de la cible. Cet algorithme trouve une bonne fonction $x \rightarrow y$

L'apprentissage statistique (supervisé)

Ensemble de données d'entraînement (*training set*)

Une **entrée** est généralement représentée par un vecteur de dimension d .

$$x \in \mathbb{R}^d$$

1	entrée, observation, <i>input</i> , x_1	cible, <i>target</i> , sortie désirée, y_1
2	entrée, observation, <i>input</i> , x_2	cible, <i>target</i> , sortie désirée, y_2
3	entrée, observation, <i>input</i> , x_3	cible, <i>target</i> , sortie désirée, y_3
⋮	<i>etc...</i>	⋮
n	entrée, observation, <i>input</i> , x_n	cible, <i>target</i> , sortie désirée, y_n

point de test



???

taille de l'ensemble, nombre d'exemples, d'échantillons.

On cherche un algorithme qui produit une sortie (*output*) qui est une **bonne prédiction** de la cible. Cet algorithme trouve une bonne fonction $x \rightarrow y$

L'apprentissage statistique (supervisé)

Ensemble de données d'entraînement (*training set*)

Une **entrée** est généralement représentée par un vecteur de dimension d .

$x \in \mathbb{R}^d$
dimensionnalité de l'entrée

1	entrée, observation, <i>input</i> , x_1	cible, <i>target</i> , sortie désirée, y_1
2	entrée, observation, <i>input</i> , x_2	cible, <i>target</i> , sortie désirée, y_2
3	entrée, observation, <i>input</i> , x_3	cible, <i>target</i> , sortie désirée, y_3
⋮	<i>etc...</i>	⋮
n	entrée, observation, <i>input</i> , x_n	cible, <i>target</i> , sortie désirée, y_n

point de test



???

taille de l'ensemble, nombre d'exemples, d'échantillons.

On cherche un algorithme qui produit une sortie (*output*) qui est une **bonne prédiction** de la cible. Cet algorithme trouve une bonne fonction $x \rightarrow y$

La tâche

prédire t à partir de x

Apprendre une fonction paramétrisée f_{θ} qui minimise l'erreur de prédiction

input target

$x \in \mathbb{R}^d$ t

n examples

x_1	x_2	x_3	x_4	x_5	t
0.32	-0.27	+1	0	0.82	113
-0.12	0.42	-1	1	0.22	34
0.06	0.35	-1	1	-0.37	56
0.91	-0.72	+1	0	-0.63	77
...

Training Set D_n

La tâche

prédire t à partir de x

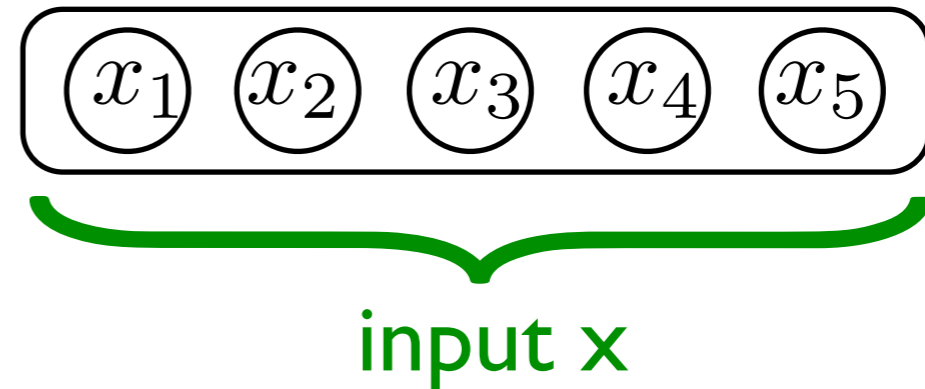
Apprendre une fonction paramétrisée f_θ qui minimise l'erreur de prédiction

input $x \in \mathbb{R}^d$ target t

n examples

x_1	x_2	x_3	x_4	x_5	t
0.32	-0.27	+1	0	0.82	113
-0.12	0.42	-1	1	0.22	34
0.06	0.35	-1	1	-0.37	56
0.91	-0.72	+1	0	-0.63	77
...

Training Set D_n



La tâche

prédire t à partir de x

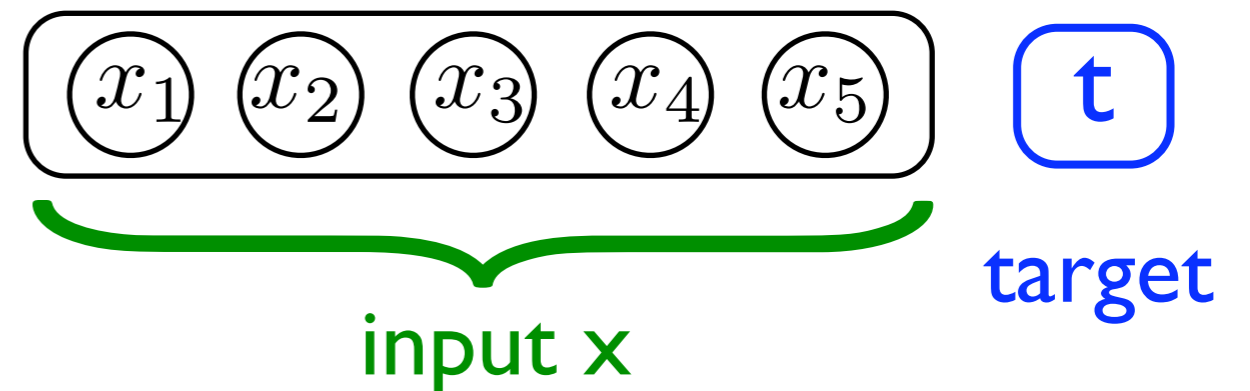
Apprendre une fonction paramétrisée f_θ qui minimise l'erreur de prédiction

input $x \in \mathbb{R}^d$ target t

n examples

x_1	x_2	x_3	x_4	x_5	t
0.32	-0.27	+1	0	0.82	113
-0.12	0.42	-1	1	0.22	34
0.06	0.35	-1	1	-0.37	56
0.91	-0.72	+1	0	-0.63	77
...

Training Set D_n



La tâche

prédire t à partir de x

input target

$x \in \mathbb{R}^d$

t

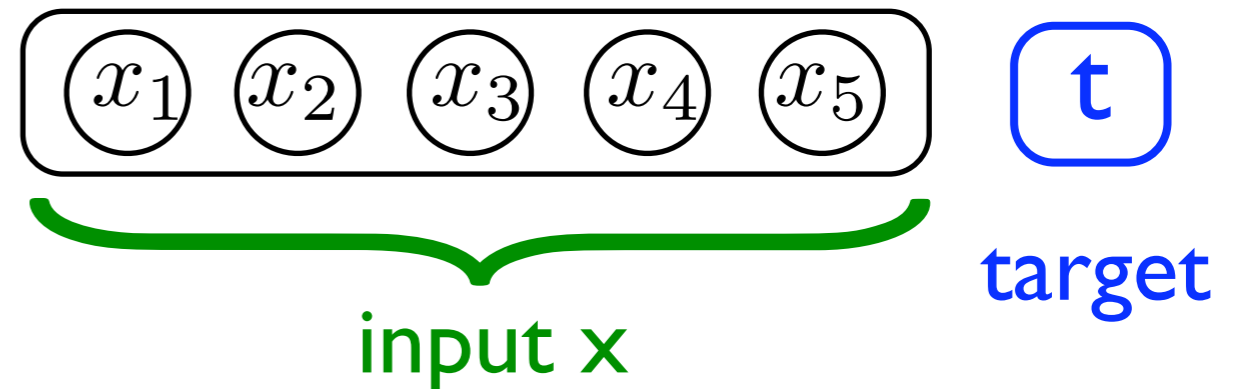
n examples

x_1	x_2	x_3	x_4	x_5	t
0.32	-0.27	+1	0	0.82	113
-0.12	0.42	-1	1	0.22	34
0.06	0.35	-1	1	-0.37	56
0.91	-0.72	+1	0	-0.63	77
...

Training Set D_n

Apprendre une fonction paramétrisée f_θ qui minimise l'erreur de prédiction

f_θ : parameters



La tâche

prédire t à partir de x

input target

$x \in \mathbb{R}^d$

t

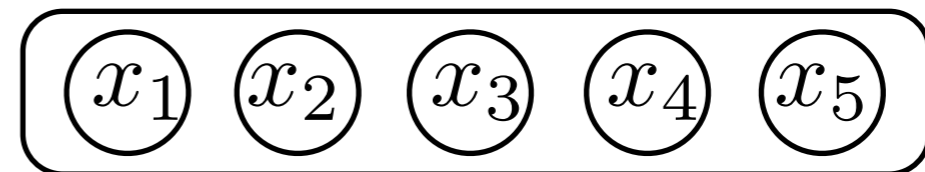
n examples

x_1	x_2	x_3	x_4	x_5	t
0.32	-0.27	+1	0	0.82	113
-0.12	0.42	-1	1	0.22	34
0.06	0.35	-1	1	-0.37	56
0.91	-0.72	+1	0	-0.63	77
...

Training Set D_n

Apprendre une fonction paramétrisée f_θ qui minimise l'erreur de prédiction

f_θ : parameters



target

input x

La tâche

prédire t à partir de x

input target

$$x \in \mathbb{R}^d$$

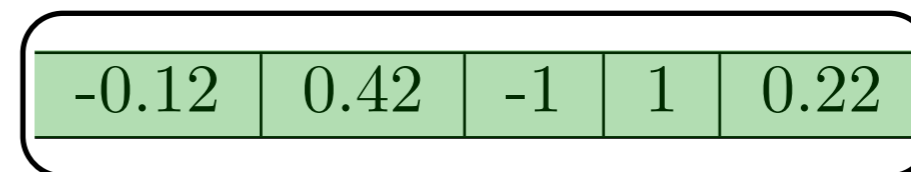
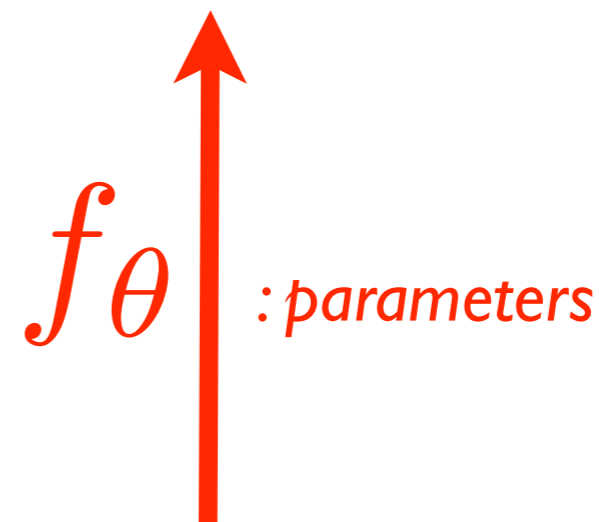
t

n examples

x_1	x_2	x_3	x_4	x_5	t
0.32	-0.27	+1	0	0.82	113
-0.12	0.42	-1	1	0.22	34
0.06	0.35	-1	1	-0.37	56
0.91	-0.72	+1	0	-0.63	77
...

Training Set D_n

Apprendre une fonction paramétrisée f_θ qui minimise l'erreur de prédiction



target

input x

La tâche

prédire t à partir de x

input target

$x \in \mathbb{R}^d$ t

n examples

x_1	x_2	x_3	x_4	x_5	t
0.32	-0.27	+1	0	0.82	113
-0.12	0.42	-1	1	0.22	34
0.06	0.35	-1	1	-0.37	56
0.91	-0.72	+1	0	-0.63	77
...

Training Set D_n

Apprendre une fonction paramétrisée f_θ qui minimise l'erreur de prédiction

output

$$y = f_\theta(x)$$

f_θ : parameters

-0.12	0.42	-1	1	0.22
-------	------	----	---	------

t

target

input x

La tâche

prédire t à partir de x

input target

$x \in \mathbb{R}^d$ t

n examples

x_1	x_2	x_3	x_4	x_5	t
0.32	-0.27	+1	0	0.82	113
-0.12	0.42	-1	1	0.22	34
0.06	0.35	-1	1	-0.37	56
0.91	-0.72	+1	0	-0.63	77
...

Training Set D_n

Apprendre une fonction paramétrisée f_θ qui minimise l'erreur de prédiction

output

$$y = f_\theta(x)$$

f_θ : parameters

-0.12	0.42	-1	1	0.22
-------	------	----	---	------

t

target

input x

La tâche

prédire t à partir de x

input target

$x \in \mathbb{R}^d$ t

n examples

x_1	x_2	x_3	x_4	x_5	t
0.32	-0.27	+1	0	0.82	113
-0.12	0.42	-1	1	0.22	34
0.06	0.35	-1	1	-0.37	56
0.91	-0.72	+1	0	-0.63	77
...

Training Set D_n

Apprendre une fonction paramétrisée f_θ qui minimise l'erreur de prédiction

output

$$y = f_\theta(x)$$

f_θ : parameters

-0.12	0.42	-1	1	0.22
-------	------	----	---	------

34

target

input x

La tâche

prédire t à partir de x

input target

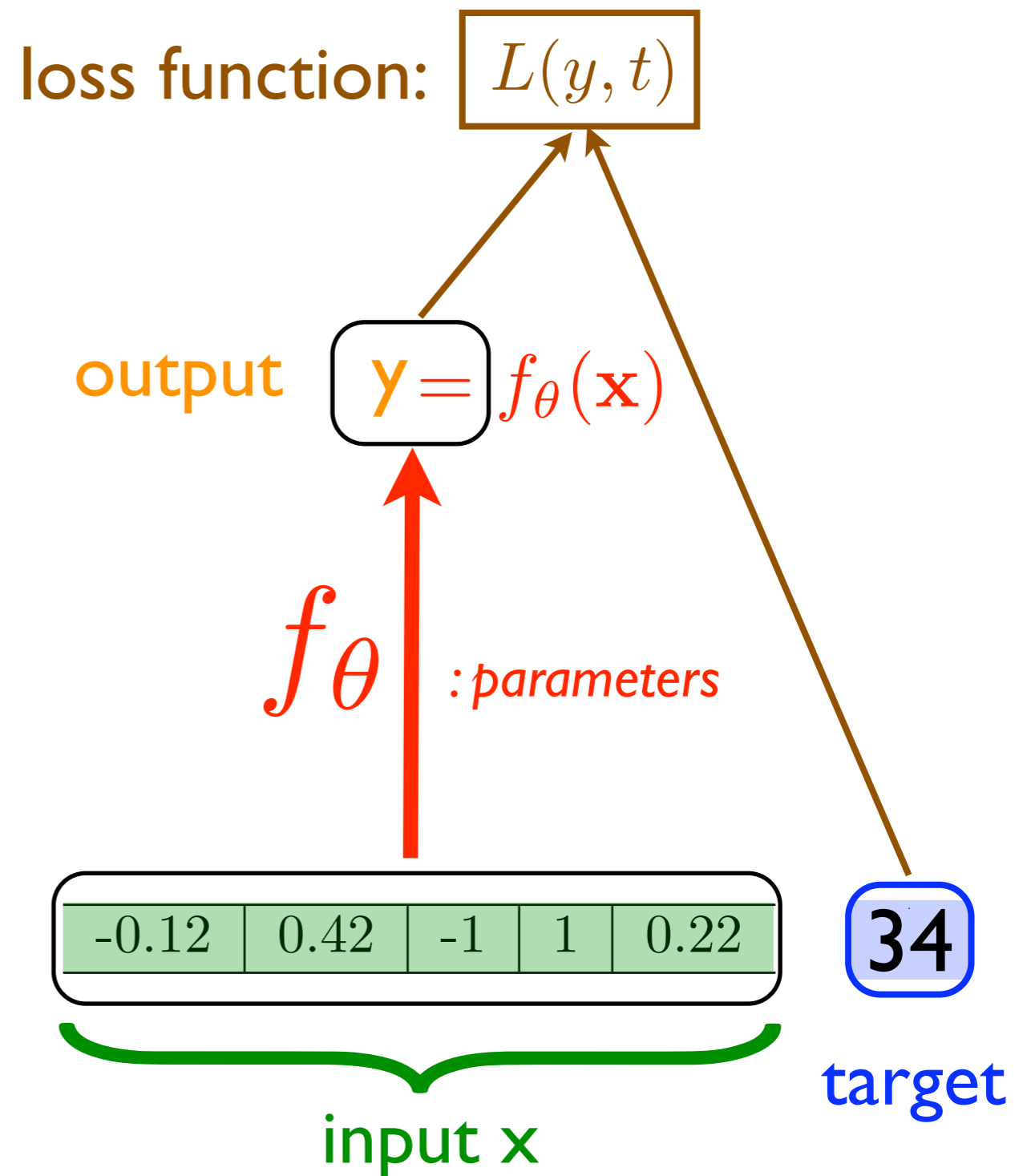
$x \in \mathbb{R}^d$ t

n examples

x_1	x_2	x_3	x_4	x_5	t
0.32	-0.27	+1	0	0.82	113
-0.12	0.42	-1	1	0.22	34
0.06	0.35	-1	1	-0.37	56
0.91	-0.72	+1	0	-0.63	77
...

Training Set D_n

Apprendre une fonction paramétrisée f_θ qui minimise l'erreur de prédiction



Les données

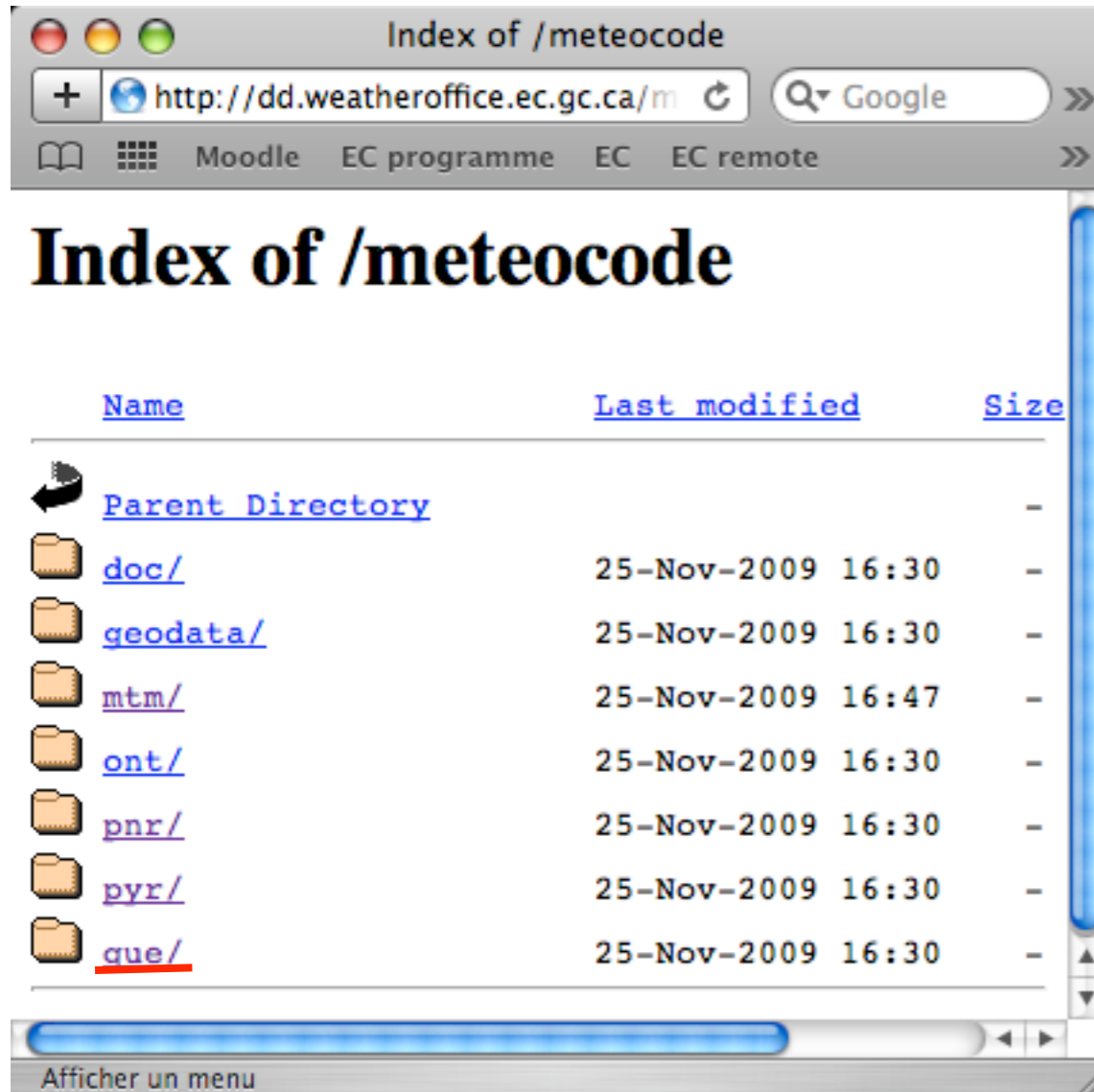
- Sont **au coeur** de l'approche d'apprentissage
- Nécessité d'un **accès à des données historiques**
- Pouvoir les **présenter** facilement dans un **format désiré** (tableau *entrée,cible*)

Données mises à notre
disposition sur le site
d'Environnement Canada.









<http://dd.weatheroffice.ec.gc.ca/meteocode/>

Données mises à notre disposition sur le site d'Environnement Canada.

<http://dd.weatheroffice.ec.gc.ca/meteocode/>



The screenshot shows a web browser window with the title "Index of /meteocode". The address bar contains the URL "http://dd.weatheroffice.ec.gc.ca/m". The browser's search bar is set to "Google". The browser's bookmark bar shows "Moodle", "EC programme", "EC", and "EC remote". The main content area displays the "Index of /meteocode" directory listing. The listing has three columns: "Name", "Last modified", and "Size". The entries are as follows:

<u>Name</u>	<u>Last modified</u>	<u>Size</u>
 Parent Directory		-
 doc/	25-Nov-2009 16:30	-
 geodata/	25-Nov-2009 16:30	-
 mtm/	25-Nov-2009 16:47	-
 ont/	25-Nov-2009 16:30	-
 pnr/	25-Nov-2009 16:30	-
 pyr/	25-Nov-2009 16:30	-
 que/	25-Nov-2009 16:30	-

At the bottom of the browser window, there is a button labeled "Afficher un menu".

Données mises à notre disposition sur le site d'Environnement Canada.

<http://dd.weatheroffice.ec.gc.ca/meteocode/>

Index of /meteocode

http://dd.weatheroffice.ec.gc.ca/meteocode/

Index of /meteocode

Name	Last modified	Size
Parent Directory		-
doc/	25-Nov-2009 16:30	-
geodata/	25-Nov-2009 16:30	-
mtm/	25-Nov-2009 16:47	-
ont/	25-Nov-2009 16:30	-
pnr/	25-Nov-2009 16:30	-
pyr/	25-Nov-2009 16:30	-
que/	25-Nov-2009 16:30	-

Afficher un menu

Index of /meteocode/que/cmml

http://dd.weatheroffice.ec.gc.ca/meteocode/que/cmml

Index of /meteocode/que/cmml

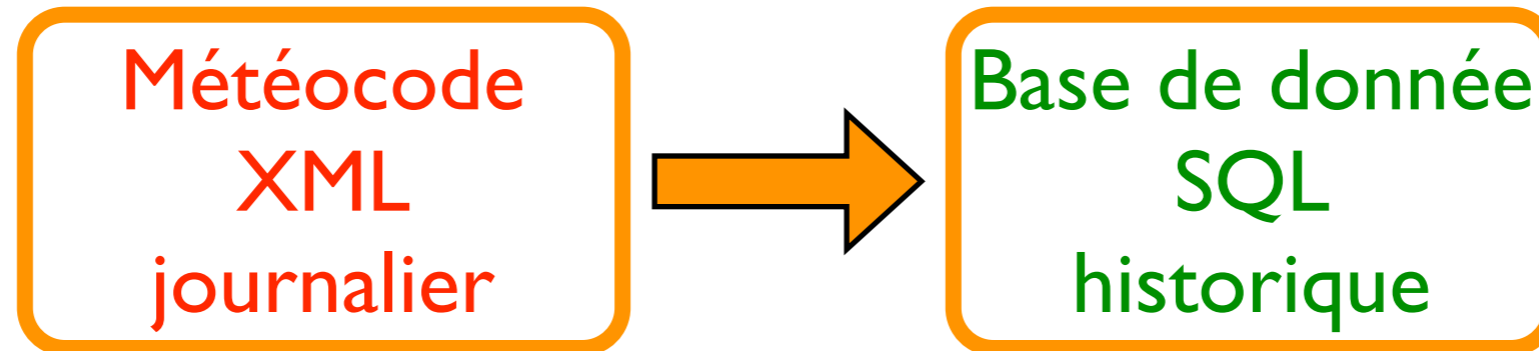
Name	Last modified	Size
Parent Directory		-
TRANSMIT.FPCN71.11.24.1226Z.AMD.xml	24-Nov-2009 12:33	463K
TRANSMIT.FPCN71.11.24.1630Z.xml	24-Nov-2009 16:43	413K
TRANSMIT.FPCN71.11.24.2045Z.xml	24-Nov-2009 20:51	478K
TRANSMIT.FPCN71.11.25.1000Z.xml	25-Nov-2009 09:51	542K
TRANSMIT.FPCN71.11.25.1630Z.xml	25-Nov-2009 17:02	468K
TRANSMIT.FPCN71.11.25.2045Z.xml	25-Nov-2009 21:08	475K
TRANSMIT.FPCN71.11.26.1000Z.xml	26-Nov-2009 09:54	508K
TRANSMIT.FPCN71.11.26.1630Z.xml	26-Nov-2009 16:29	441K
TRANSMIT.FPCN71.11.26.2045Z.xml	26-Nov-2009 20:49	244K
TRANSMIT.FPCN71.11.27.1000Z.xml	27-Nov-2009 10:40	278K
TRANSMIT.FPCN71.11.27.1423Z.AMD.xml	27-Nov-2009 14:37	264K
TRANSMIT.FPCN71.11.27.1630Z.xml	27-Nov-2009 16:27	389K
TRANSMIT.FPCN71.11.27.2045Z.xml	27-Nov-2009 21:03	427K
TRANSMIT.FPCN71.11.28.1000Z.xml	28-Nov-2009 09:56	502K
TRANSMIT.FPCN71.11.28.1630Z.xml	28-Nov-2009 16:37	426K
TRANSMIT.FPCN71.11.28.2045Z.xml	28-Nov-2009 21:05	446K
TRANSMIT.FPCN71.11.29.1000Z.xml	29-Nov-2009 10:36	524K
TRANSMIT.FPCN71.11.29.1630Z.xml	29-Nov-2009 16:45	460K
TRANSMIT.FPCN71.11.29.2045Z.xml	30-Nov-2009 01:54	492K
TRANSMIT.FPCN71.11.30.0111Z.AMD.xml	30-Nov-2009 02:01	436K

Afficher un menu

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<cmml xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:noNamespaceSchemaLocation="/data/wxodfeed/dd/public
  <head>
    <product operational-mode="developmental">
      <title>FPCN71</title>
      <field>meteorological</field>
      <category>forecast</category>
      <creation-date refresh-frequency="PT13H">2009-11-25T20:40:03Z</creation-date>
      <current-issue>2009-11-25T20:45:00Z</current-issue>
      <valid-begin-time>2009-11-25T21:00:00Z</valid-begin-time>
      <valid-end-time>2009-11-27T11:00:00Z</valid-end-time>
      <next-issue>2009-11-26T10:00:00Z</next-issue>
      <timezone>EST5EDT</timezone>
      <type>regular</type>
    </product>
    <source>
      <more-information>http://weatheroffice.ec.gc.ca</more-information>
      <production-center>Meteorological Service of Canada<sub-center>CWUL</sub-center></production-center>
      <feedback/>
      <credit/>
      <credit-logo/>
      <disclaimer>http://www.weatheroffice.ec.gc.ca/mainmenu/disclaimer_e.html</disclaimer>
    </source>
  </head>
  <data>
    <forecast>
      <meteocode-forecast>
        <location>
          <msc-zone-code>r71.1</msc-zone-code>
          <msc-zone-name lang="fr">Montréal métropolitain - Laval </msc-zone-name>
          <msc-zone-name lang="en">Metro Montréal - Laval</msc-zone-name>
        </location>
        <parameters>
          <warning-list/>
          <cloud-list units="deci">
            <cloud-cover start="2009-11-25T21:00:00Z" end="2009-11-25T21:00:00Z" ceiling-code="2">9</cloud-cover>
            <cloud-cover start="2009-11-25T22:00:00Z" end="2009-11-25T22:00:00Z" ceiling-code="2">9</cloud-cover>
            <cloud-cover start="2009-11-25T23:00:00Z" end="2009-11-25T23:00:00Z" ceiling-code="2">9</cloud-cover>
            <cloud-cover start="2009-11-26T00:00:00Z" end="2009-11-26T00:00:00Z" ceiling-code="2">9</cloud-cover>
            <cloud-cover start="2009-11-26T01:00:00Z" end="2009-11-26T01:00:00Z" ceiling-code="2">9</cloud-cover>
            <cloud-cover start="2009-11-26T02:00:00Z" end="2009-11-26T02:00:00Z" ceiling-code="2">9</cloud-cover>
            <cloud-cover start="2009-11-26T03:00:00Z" end="2009-11-26T03:00:00Z" ceiling-code="2">9</cloud-cover>
            <cloud-cover start="2009-11-26T04:00:00Z" end="2009-11-26T04:00:00Z" ceiling-code="2">9</cloud-cover>
          </cloud-list>
        </parameters>
      </meteocode-forecast>
    </forecast>
  </data>
</cmml>
```

Fichier météo code
au format XML

Premier objectif



- **Comprendre** les données, leur format XML, et leur signification.
- Mettre en place un processus de **collecte automatique** de ces données.
- Peupler une **base de donnée relationnelle** (SQL) avec les **données historiques**.
- Ceci permettra des requêtes efficaces pour obtenir des exemples d'apprentissage.



navigateur d'objets

- ▼ Serveurs (1)
 - ▼ langevse_db (gershwin.iro.umontreal)
 - ▼ Bases de données (22)
 - ❌ bastienf_db
 - ❌ bengioy_db
 - ❌ bergstrj_db
 - ❌ breuleuo_db
 - ❌ desjagui_db
 - ❌ erhandum_db
 - ❌ glorotxa_db
 - ❌ lamblinp_db
 - ▼ langevse_db
 - ▶ Catalogues (2)
 - ▼ Schémas (1)
 - ▼ public
 - 🏠 Domaines (0)
 - 🔧 Configurations FTS (0)
 - 📖 Dictionnaires FTS (0)
 - 📊 Analyseurs FTS (0)
 - 📄 Modèles FTS (0)
 - 👤 Fonctions (0)
 - ▶ Séquences (1)
 - ▼ Tables (28)
 - ▼ accum_list
 - ▶ Colonnes (8)
 - ▶ Contraintes (3)
 - 📄 Index (0)
 - 📄 Règles (0)
 - 🔗 triggers (0)

Propriétés Statistiques Dépendances Objets dépendants

Nom de table	Lignes insérées	Lignes mises à jour	Lignes effacées	Lignes
accum_list	92007	0	0	0
cloud_exceptions_list	3	0	0	0
cloud_list	40457	0	0	0
encode_list	0	0	0	0
freezing_spray_exceptions_list	0	0	0	0
freezing_spray_list	0	0	0	0
ice_cover_list	0	0	0	0
meta_info	550	0	20	0
msc_zone_list	66	66	0	66
precipitation_exceptions_list	49	0	0	0
precipitation_list	22815	0	0	0
probability_of_precipitation_list	31357	0	0	0
snow_level_list	0	0	0	0
temperature_air_exceptions_list	2	0	0	0
temperature_air_list	176242	0	0	0
temperature_climatology_exceptions_list	0	0	0	0
temperature_climatology_list	47301	0	0	0
temperature_dew_point_exceptions_list	0	0	0	0
temperature_dew_point_list	176242	0	0	0
temperature_sea_surface_exceptions_list	0	0	0	0
temperature_sea_surface_list	0	0	0	0
uv_index_list	3630	0	0	0
visibility_exceptions_list	3	0	0	0
visibility_list	1593	0	0	0
warning_list	230	0	0	0
wave_height_list	0	0	0	0
wind_exceptions_list	19	0	0	0
wind_list	52219	0	0	0

Pas de limite

	title [PK] caracte	msc_zone [PK] chara	current_issue [PK] timestamp without tim	start [PK] timestamp without ti	end [PK] timestamp without ti	type charact	lower_limit real	upper_limit real
1	FPCN71	r71.1	2009-11-21 16:30:00	2009-11-21 17:00:00	2009-11-23 05:00:00		0	0
2	FPCN71	r71.1	2009-11-21 20:45:00	2009-11-21 21:00:00	2009-11-23 11:00:00		0	0
3	FPCN71	r71.1	2009-11-22 10:00:00	2009-11-22 10:00:00	2009-11-24 05:00:00		0	0
4	FPCN71	r71.1	2009-11-22 12:14:00	2009-11-22 12:00:00	2009-11-24 05:00:00		0	0
5	FPCN71	r71.1	2009-11-22 16:30:00	2009-11-22 17:00:00	2009-11-24 05:00:00		0	0
6	FPCN71	r71.1	2009-11-22 20:45:00	2009-11-22 21:00:00	2009-11-24 11:00:00		0	0
7	FPCN71	r71.1	2009-11-23 01:51:00	2009-11-23 01:00:00	2009-11-24 11:00:00		0	0
8	FPCN71	r71.1	2009-11-23 10:00:00	2009-11-23 10:00:00	2009-11-25 05:00:00		0	0
9	FPCN71	r71.1	2009-11-23 10:02:00	2009-11-23 10:00:00	2009-11-25 05:00:00		0	0
10	FPCN71	r71.1	2009-11-23 16:30:00	2009-11-23 17:00:00	2009-11-25 05:00:00		0	0
11	FPCN71	r71.1	2009-11-23 20:45:00	2009-11-23 21:00:00	2009-11-25 11:00:00		0	0
12	FPCN71	r71.1	2009-11-24 00:43:00	2009-11-24 00:00:00	2009-11-25 11:00:00		0	0
13	FPCN71	r71.1	2009-11-24 10:00:00	2009-11-24 10:00:00	2009-11-26 05:00:00		0	0
14	FPCN71	r71.1	2009-11-24 16:30:00	2009-11-24 17:00:00	2009-11-26 05:00:00		0	0
15	FPCN71	r71.1	2009-11-24 20:45:00	2009-11-24 21:00:00	2009-11-25 23:00:00		0	0
16	FPCN71	r71.1	2009-11-24 20:45:00	2009-11-25 23:00:00	2009-11-26 00:00:00	rain	0.33	0.67
17	FPCN71	r71.1	2009-11-24 20:45:00	2009-11-26 00:00:00	2009-11-26 01:00:00	rain	0.33	0.67
18	FPCN71	r71.1	2009-11-24 20:45:00	2009-11-26 01:00:00	2009-11-26 02:00:00	rain	0.33	0.67
19	FPCN71	r71.1	2009-11-24 20:45:00	2009-11-26 02:00:00	2009-11-26 03:00:00	rain	1.67	1.67
20	FPCN71	r71.1	2009-11-24 20:45:00	2009-11-26 03:00:00	2009-11-26 04:00:00	rain	1.67	1.67
21	FPCN71	r71.1	2009-11-24 20:45:00	2009-11-26 04:00:00	2009-11-26 05:00:00	rain	1.67	1.67
22	FPCN71	r71.1	2009-11-24 20:45:00	2009-11-26 05:00:00	2009-11-26 06:00:00	rain	0	0.33
23	FPCN71	r71.1	2009-11-24 20:45:00	2009-11-26 06:00:00	2009-11-26 07:00:00	rain	0	0.33

Bloc notes

Bizarries constatées

- Un nouveau fichier XML (non .AMD) apparaît parfois 2 à 3mn aux côtés du premier, avec une ou deux données modifiées. (régulé?)
- Régions différentes avec zone-code identique.

Exemple:

[.../pnr/cmml/TRANSMIT.FPNT14.06.26.15Z.1430.xml](#)

title : FPNT14
zone-code : **r23**
zone-name : **Kugaaruk**

[.../pyr/cmml/TRANSMIT.FPVR16.06.24.18Z.1734.xml](#)

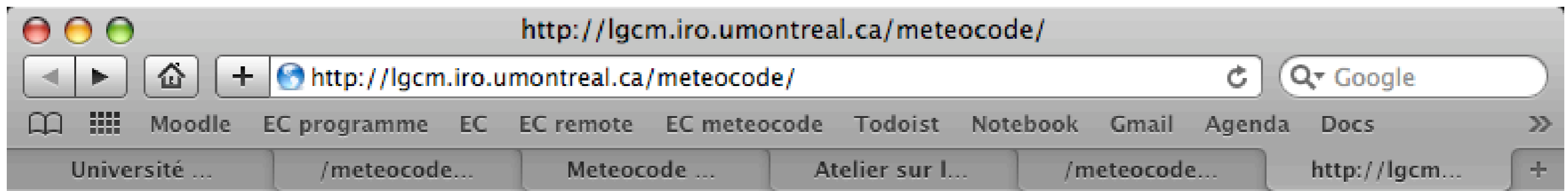
title : FPVR16
zone-code : **r23**
zone-name : **Williston**

Les deux titres contiennent d'autres régions ne semblant pas être maritimes (Kugaaruk est une ville, et bien que Williston soit un lac au BC, le titre FPVR contient d'autres régions telles que rv7: Metro Vancouver ou rv12: Whistler)

Deuxième objectif

Permettre un accès interactif via le web à la base de données ainsi constituée.

(pour exploration et visualisation simple)



Meteocode data access project

(Test version)

Sébastien Langevin

Université de Montréal

2009 August 26th

This website allows users to query for raw meteorological forecast data produced by Environment Canada
<http://dd.weatheroffice.gc.ca>.

Note that multiple forecasts normally cover the same period of time. Thus two modes of access are available:

For a given **RANGE** of time, obtain forecasts with a specified lead time (forecasts with lead times closest but greater to the specified one will be returned.)

Time range query

For a given **INSTANT** of time, obtain all available forecasts.

Punctual query

MSC-zone-code chooser

http://lgcm.iro.umontreal.ca/meteocode/cgi-bin/timerangeview.cgi?msc_zone_ Google

Moodle EC programme EC EC remote EC meteocode Todoist Notebook Gmail Agenda Docs Bookmarks

Université d... /meteocode... Meteocode t... Atelier sur l... /meteocode... MSC-zone-...

Zone code	Zone name(fr)	Zone name(en)
r71.1	Montréal métropolitain - Laval	Metro Montréal - Laval
r71.10	Haute-Gatineau - Lièvre - Papineau	Upper Gatineau - Lièvre - Papineau
r71.11	Pontiac	Pontiac
r71.12	Abitibi	Abitibi
r71.13	Témiscamingue	Témiscamingue
r71.14	Parent - réservoir Gouin	Parent - Gouin Reservoir
r71.15	Matagami	Matagami
r71.16	Waskaganish	Waskaganish
r71.2	Vaudreuil - Soulanges - Huntingdon	Vaudreuil - Soulanges - Huntingdon
r71.3	Vallée du Richelieu - Saint-Hyacinthe	Richelieu Valley - Saint-Hyacinthe
r71.4	Lanaudière	Lanaudière
r71.5	Lachute - Saint-Jérôme	Lachute - Saint-Jérôme
r71.6	Laurentides	Laurentians
r71.7	Parc du Mont-Tremblant - Saint-Michel-des-Saints	Mont-Tremblant Park - Saint-Michel-des-Saints
r71.8	Mont-Laurier	Mont-Laurier
r71.9	Réserve faunique La Vérendrye	La Vérendrye Wildlife Reserve

Zone-code:

Start Date (YYYY-MM-DD HH):

Timespan:

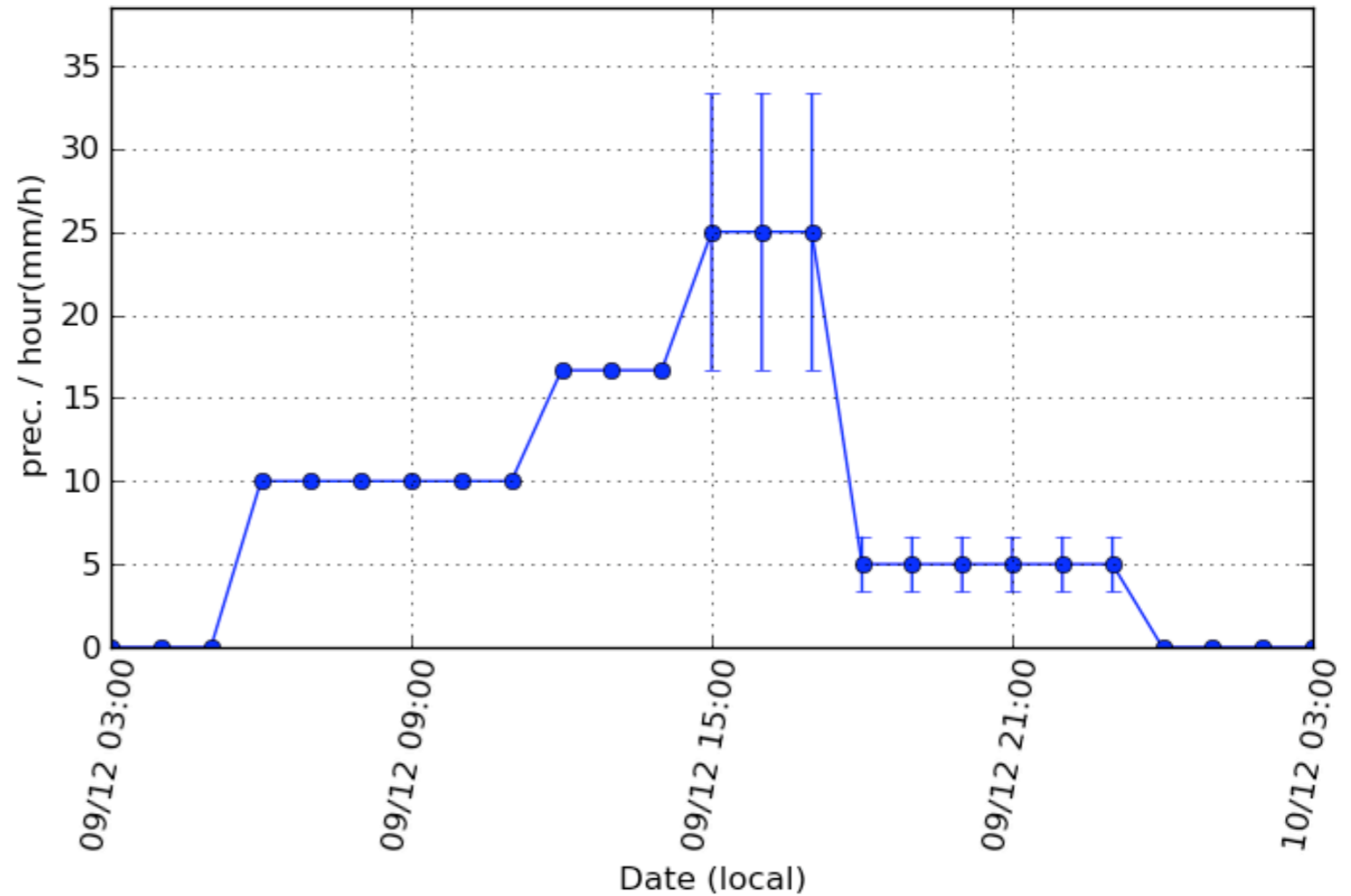
Lead time:

Time zone:

- ✓ Accumulations/hour (rain/snow)
- Cloud cover
- Ceiling height
- Probability of precipitation
- Temperature (air)
- Temperature (dew-point)
- Temperature (climatology)
- Wind speed
- Wind speed and direction
- UV index

Precipitations (quantity)

Region r71.1, Lead time > 0 hours



Temperature (air)
Region: r71.1, Lead time > 0 hours

Bulletin_date	Prediction_date	lower_limit_c	upper_limit_c
2009-12-08 05:00:00-05:00	2009-12-09 04:00:00-05:00	-5.0	-5.0
2009-12-07 16:00:00-05:00	2009-12-09 06:00:00-05:00	-5.0	-5.0
2009-12-08 05:00:00-05:00	2009-12-09 07:00:00-05:00	-5.0	-5.0
2009-12-08 05:00:00-05:00	2009-12-09 10:00:00-05:00	-4.0	-4.0
2009-12-08 05:00:00-05:00	2009-12-09 13:00:00-05:00	-2.0	-2.0
2009-12-08 05:00:00-05:00	2009-12-09 16:00:00-05:00	0.0	0.0
2009-12-08 05:00:00-05:00	2009-12-09 19:00:00-05:00	0.0	0.0
2009-12-04 16:00:00-05:00	2009-12-09 20:00:00-05:00	3.0	3.0
2009-12-08 05:00:00-05:00	2009-12-09 22:00:00-05:00	0.0	0.0
2009-12-08 05:00:00-05:00	2009-12-10 00:00:00-05:00	0.0	0.0

Display format:

Accès à toute l'historique

Zone-code:

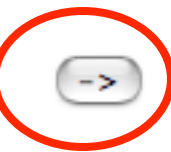
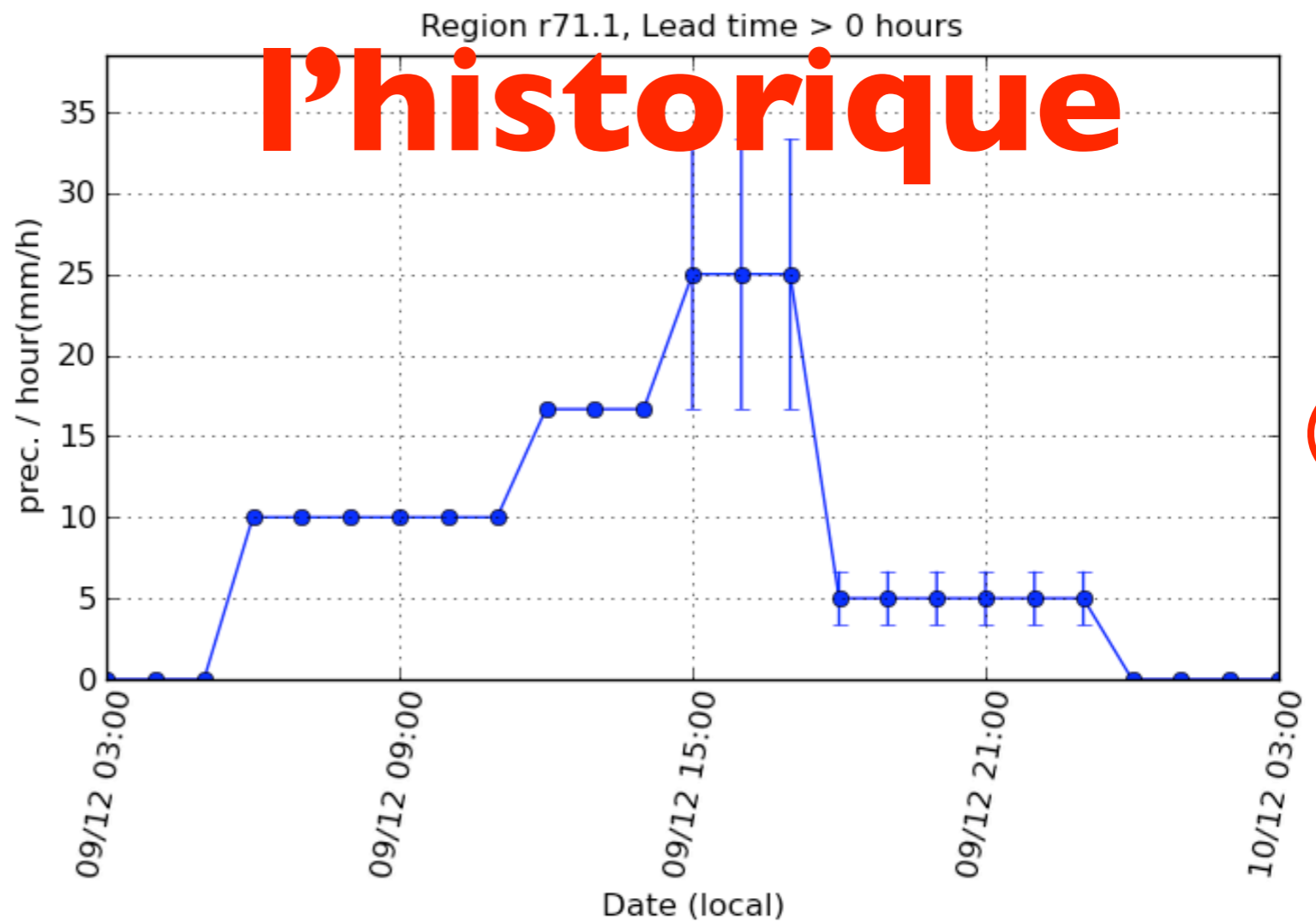
Start Date (YYYY-MM-DD HH)

Timespan: day(s)

Lead time: hour(s)

Time zone:

- Accumulations/hour (rain/snow)
- Cloud cover
- Ceiling height
- Probability of precipitation
- Temperature (air)
- Temperature (dew-point)
- Temperature (climatology)
- Wind speed
- Wind speed and direction
- UV index



Temperature (air)
Region: r71.1, Lead time > 0 hours

Bulletin_date	Prediction_date	lower_limit_c	upper_limit_c
2009-12-08 05:00:00-05:00	2009-12-09 04:00:00-05:00	-5.0	-5.0
2009-12-07 16:00:00-05:00	2009-12-09 06:00:00-05:00	-5.0	-5.0
2009-12-08 05:00:00-05:00	2009-12-09 07:00:00-05:00	-5.0	-5.0
2009-12-08 05:00:00-05:00	2009-12-09 10:00:00-05:00	-4.0	-4.0
2009-12-08 05:00:00-05:00	2009-12-09 13:00:00-05:00	-2.0	-2.0
2009-12-08 05:00:00-05:00	2009-12-09 16:00:00-05:00	0.0	0.0
2009-12-08 05:00:00-05:00	2009-12-09 19:00:00-05:00	0.0	0.0
2009-12-04 16:00:00-05:00	2009-12-09 20:00:00-05:00	3.0	3.0
2009-12-08 05:00:00-05:00	2009-12-09 22:00:00-05:00	0.0	0.0
2009-12-08 05:00:00-05:00	2009-12-10 00:00:00-05:00	0.0	0.0

Display format:



Bilan

- **Leçon apprise:** quantité de données impressionnante à gérer!
- Nous savons que ce n'est que la **partie émergée de l'iceberg...**
- Les **choses sont en places** pour collecter une base de données historiques (la *matière première* pour tout apprentissage)
- **Et après...**
Que peut-on envisager faire à partir de là?

Une possibilité?



- Un particulier peut obtenir des prédictions de haute qualité pour un zone-code.
- Mais un **zone-code** peut représenter une **zone étendue**.
- Idéalement: fournir des prédictions affinées à une **coordonnée géographique (x,y,z,t) très précise**.
(ex. au dessus d'un champ).

Une possibilité?

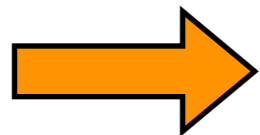


- Un particulier peut obtenir des prédictions de haute qualité pour un zone-code.
- Mais un **zone-code** peut représenter une **zone étendue**.
- Idéalement: fournir des prédictions affinées à une **coordonnée géographique (x,y,z,t) très précise**.
(ex. au dessus d'un champ).



Technique possible

- **A partir** des prévisions des **zone-code** seulement.
(c'est tout ce que nous avons pour l'instant!)



Interpolation intelligente

Ex: Régression de Processus Gaussien (~Kriging)

coordonnée d'intérêt: $u = (x, y, z, t, \dots, context)$

variable météorologique d'intérêt: $f(u) \sim \mathcal{N}(\mu(u), \sigma^2(u))$

forme de la solution: $\mu(u) = \sum_{i=1}^n \alpha_i \underbrace{K(u, u_i)}_{\substack{\text{noyau} \\ \text{ou fonction de} \\ \text{covariance}}}$

Possibilité d'**apprendre les paramètres du noyau** à partir des données historiques.

Ou avec plus d'information...

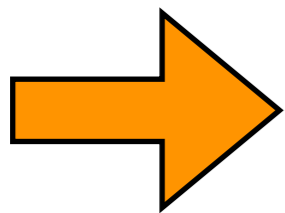
Ce dont **EC** dispose:

- Prédiction de très **haute qualité** (corrigées à la main par météorologue) **à faible résolution** (zone code).
- Prédiction **brutes du modèle** à **très haute résolution**.

Ou avec plus d'information...

Ce dont **EC** dispose:

- Prédiction de très **haute qualité** (corrigées à la main par météorologue) **à faible résolution** (zone code).
- Prédiction **brutes du modèle** à **très haute résolution**.

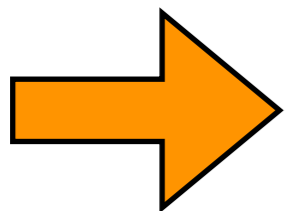


Possibilité de combiner les deux:
apprendre à “interpoler” les **corrections** des experts, sur la grille haute résolution.

Ou avec plus d'information...

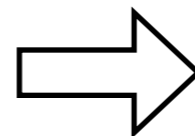
Ce dont **EC** dispose:

- Prédiction de très **haute qualité** (corrigées à la main par météorologue) **à faible résolution** (zone code).
- Prédiction **brutes du modèle** à **très haute résolution**.



Possibilité de combiner les deux:
apprendre à “interpoler” les **corrections** des experts, sur la grille haute résolution.

Différents algorithmes d'apprentissage envisageable (processus Gaussiens, réseaux de neurones, etc...)



Recherche!



Questions