Classification methods

(CIV6540 - Machine Learning for Civil Engineers)

Professor: James-A. Goulet

Département des génies civil, géologique et des mines
Polytechnique Montréal

Probabilistic Machine Learning for Civil Engineers.


What is classification?

Data

\[ D = \{(x_i, y_i), \forall i = 1 : D\} \]

- Covariate attribute
- regressor

\[ x_i \in \mathbb{R} \]

\[ y_i \in \{-1, 1\} : \text{Observation} \]

**Classification methods:** mathematical models for \( \Pr(Y|x) \)
2 Types of classification methods

There are two types of classification methods:

1. **Generative**

   \[
   p(y|x) = \frac{p(x|y)p(y)}{p(x)}
   \]

2. **Discriminative**

   \[
   p(y|x) = g(x)
   \]
Module #6 Outline

Topics organization

Introduction
Generative classifier
Logistic Regression
Gaussian Process Classification (GPC)
NN
Summary

What is classification?

Probability theory
1. Revision probability & linear algebra
2. Probability distributions

Machine Learning Basics
0. Introduction
3. Bayesian Estimation
  \[ p(A|B) = \frac{p(B|A)p(A)}{p(B)} \]
4. MCMC sampling & Newton

Supervised learning
5. Regression
6. Classification
7. State-space model for time-series

Decision Making
8. Decision Theory
9. AI & Sequential decision problems

Revision probability & linear algebra

Probability distributions

Bayesian Estimation

MCMC sampling & Newton

Regression

Classification

State-space model for time-series

Decision Theory

AI & Sequential decision problems
Section Outline

Generative classifier

2.1 Introduction
2.2 Formulation
2.3 Multiple classes
2.4 Example
2.5 Classic Generative Methodologies
2.6 Summary
Classification and Bayes rule

Data
\[ \mathcal{D} = \{(x_i, y_i), \forall i = 1 : D\} \]
\[ x_i \in \mathbb{R} : \left\{ \begin{array}{l} \text{Covariate} \\ \text{attribute} \\ \text{regressor} \end{array} \right. \]
\[ y_i \in \{-1, 1\} : \text{Observation} \]

\[ f(y, x) = p(y|x)f(x) = f(x|y)p(y) \rightarrow p(y|x) = \frac{f(x|y) \cdot p(y)}{f(x)} \]

\[ \text{Pr}(y = i|x) = \frac{f(x|y = i) \cdot p(y = i)}{f(x) \sum_j f(x|y = j) \cdot p(y = j)} \]
Formulation - Binary classification

Posterior probability of a class

\[ p(y|x) = \begin{cases} \Pr(Y = -1|x) \\ \Pr(Y = 1|x) = 1 - \Pr(Y = -1|x) \end{cases} \]

Prior probability of a class

\[ p(y) = \begin{cases} \Pr(Y = -1) \\ \Pr(Y = 1) \end{cases} \]

Likelihood of \( x_i \) given a class

\[ f(x|y) = \begin{cases} f(X = x|y = -1) \\ f(X = x|y = 1) \end{cases} \]

Normalization constant

\[ f(x) = \sum_{j} f(x|y = j) \cdot p(y = j) = \text{cte.} \]
Learning parameters - Binary classification

\[
p(y|x) = \frac{\text{likelihood} \cdot \text{prior}}{\text{norm. cte.}}
\]

We need to learn the PDF/PMF parameters from \( D \)

MLE/Bayes
Formulation - Special case $f(x|y) = \mathcal{N}(x; \mu, \sigma^2)$

If we assume that $f(x|y = j) = \mathcal{N}(x; \mu_{x|y_j}, \sigma^2_{x|y_j})$, and that the number of available data $D$ is large

MLE approximation of $\mu_{x|y_j}$ & $\sigma^2_{x|y_j}$

$$\hat{\mu}_{x|y_j} = \frac{1}{\#\{i: y_i = j\}} \sum_i x_i, \ \forall \{i: y_i = j\}$$

$$\hat{\sigma}^2_{x|y_j} = \frac{1}{\#\{i: y_i = j\} - 1} \sum_i (x_i - \hat{\mu}_{x|y_j})^2, \ \forall \{i: y_i = j\}$$

MLE approximation of $p(y)$

$$\hat{p}(y) = \left\{ \hat{P}r(y = j) = \frac{\#\{i: y_i = j\}}{D} \right\}$$
Example – Binary classification

\[ p(y = +1|x) \]

\[ f(x|y = -1), f(x|y = +1) \]

Formulation

\[
\begin{align*}
\{x_i, y_i = -1\} & \quad \{x_i, y_i = +1\} \\
\{x, y = -1\} & \quad \{x, y = +1\}
\end{align*}
\]
Multiattribute classification

What if there is **more than one attribute** \( x = [x_1, x_2, \cdots, x_X] \)?

**Employ joint PDFs:** \( f(x|y = j) \)

\[
f(x|y = j) = \mathcal{N}(x; \mu_{x|y_j}, \Sigma_{x|y_j})
\]

MLE/Bayes
Example – Multiattribute classification

\[ p(Y = 1|x) \]

\[ \{x_i, y_i = 0\} \quad \{x_i, y_i = 1\} \]

\[ f(X|y) = \begin{cases} \text{blue} & \text{if } y = 0 \\ \text{red} & \text{if } y = 1 \end{cases} \]
Formulation - Multi-classes classification

Posterior probability of a class

\[ p(y|x) = \{ \Pr(y = j|x) \} \]

Prior probability of a class

\[ p(y) = \{ \Pr(y = j) \} \]

Likelihood of \( x_i \) given a class

\[ f(x|y) = \{ f(X = x|y = i) \} \]

Normalization constant

\[ f(x) = \sum_j f(x|y = j) \cdot p(y = j) \]
Context

**Societal challenge**
Right after an earthquake
– which building is safe?
– which one should be evacuated?

**Solution**
Monitor structures

**Scientific challenges**
Real-time data-driven learning
**Frequency ratio**: $x = \frac{f_{q_{\text{post-earthquake}}}}{f_{q_{\text{pre-earthquake}}}}$

- Dunand et al., 2004 (Algeria)
- Omori, 1922 (Japan)
- Régnier et al., 2013 (Martinique)
- Vidal et al., 2013 (Spain)
- Mucciarelli et al., 2004 (Italy)
- Clinton et al., 2006 (USA)

**Upper bound**
Frequency ratio: \( x = \frac{f_{q_{\text{post-earthquake}}}}{f_{q_{\text{pre-earthquake}}}} \)
Frequency ratio: \( x = \frac{f_{q\cdot\text{post-earthquake}}}{f_{q\cdot\text{pre-earthquake}}} \)
Frequency ratio: \( x = \frac{f_{\text{post-earthquake}}}{f_{\text{pre-earthquake}}} \)

We need to learn as inspections are realized after an earthquake
Step 1 - Probabilistic learning

**Given:** \( D = \{ x_i, d_i \}, \forall i = 1 : D \)
\[\in (0,1) \in \{0:5\}\]

**Goal:** Learn what is the conditional PDF (predictive) of \( X \) given \( D \)

\[
f_{X \mid D}(x \mid d) = \mathcal{B}(x; \mu(d), \sigma(d))
\]

\[
f_{\theta \mid X}(\theta \mid D) = \frac{f_{X \theta}(D \mid \theta) \cdot f_{\theta}(\theta)}{f_X(D)}
\]

cte.

\[
f_{X \mid D, D}(x \mid d, D) = \int f_{X \mid D}(x \mid d) f_{\theta \mid X}(\theta \mid D) d\theta
\]
Step 1 - Probabilistic learning (likelihood & prior)

\[
\underbrace{f_{\theta|X}(\mu(d), \sigma(d)|D)}_{\text{posterior}} \propto \underbrace{f_{X|\theta}(D|\mu(d), \sigma(d))}_{\text{likelihood}} \cdot \underbrace{f_{\theta}(\mu(d), \sigma(d))}_{\text{prior}}
\]

\[
f_{X|\mu(d),\sigma(d)}(D|\theta) = \prod_{\{i:d_i=d\}} \mathcal{L}(x_i|\mu(d), \sigma(d))
\]

\[
\mathcal{L}(x_i|\mu(d), \sigma(d)) = \begin{cases} 
  f_{X|\theta}(x_i|\mu(d), \sigma(d)), & \text{if } x_i \text{ direct obs.} \\
  F_{X|\theta}(x_i|\mu(d), \sigma(d)), & \text{if } x_i \text{ upper bound obs.} 
\end{cases}
\]

We assume no prior information except for the constraint that

\[
\mu(0) > \mu(1) > \cdots > \mu(5)
\]
Application - \( f_{\theta|X}(\mu(d), \sigma(d)|D) \) & \( f_{X|D}(x|d) \)
Application - \( f_{\theta|X}(\mu(d), \sigma(d)|D) \) & \( f_{X|D}(x|d) \)

\[
f(x|d, D) = \int \int \mathcal{B}(x; \mu(d), \sigma(d)) f_{\theta|X}(\mu(d), \sigma(d)|D) d\mu(d) d\sigma(d)
\]

We have \( f(x|d, D) \), we want \( p(d|x, D) \)
Step 2- Probabilistic prognosis

\[
p(d|x, D) = \int \frac{f(x; \theta(d)) \cdot p(d)}{\sum_{d'=0}^{5} f(x; \theta(d')) \cdot p(d')} \cdot f(\theta(d)|D) d\theta
\]
Post-earthquake damage simulation for San Francisco

389 buildings ≥ 10 story
Example

Professor: J-A. Goulet

Classification methods | V1.4 | Machine Learning for Civil Engineers
**Example**

**Classification methods**

- Generative classifier
- Logistic Regression
- Gaussian Process Classification (GPC)
- NN
- Summary

---

**Diagram Description**

- **Top Graph:**
  - Functions $f_{x|D}(x|d)$ for $d=0$ to $d=5$.
  - Different colors represent different conditions.

- **Middle Graph:**
  - Graph showing $P(D \in S|x)$.
  - Conditions are color-coded:
    - Red: $S=\{4,5\}$ (unsafe)
    - Orange: $S=\{2,3\}$ (unsafe*)
    - Green: $S=\{0,1\}$ (safe)

- **Bottom Graph:**
  - Graph showing the number of inspections performed.
  - False +: 0% | False -: 0%
  - Unknown condition
  - Predicted condition
  - Observed condition

---

**Legend:**

- Unknown condition
- Predicted condition
- Observed condition

---

*Polytechnique Montréal*
Introduction

Generative classifier

Logistic Regression

Gaussian Process Classification (GPC)

NN

Summary

Example

Professor: J-A. Goulet

Classification methods | V1.4 | Machine Learning for Civil Engineers
Example

- **Generative classifier**
  - Logistic Regression
  - Gaussian Process Classification (GPC)

- **Summary**

**Classification methods**

**V1.4** | Machine Learning for Civil Engineers

---

*Polytechnique Montréal*
Example

\( f_{\mathcal{X} \mid D}(x \mid d) \)

- \( d=0 \)
- \( d=1 \)
- \( d=2 \)
- \( d=3 \)
- \( d=4 \)
- \( d=5 \)

Probability of condition being in one of the sets:

- \( S=\{4,5\} \) (unsafe)
- \( S=\{2,3\} \) (unsafe*)
- \( S=\{0,1\} \) (safe)

False +: 0% | False -: 0%

Number of inspections performed [%]

Unknown condition
Predicted condition
Observed condition
Example

**Graph 1:**
- **Axes:**
  - x-axis: \( x \)
  - y-axis: \( f_{X|D}(x|d) \)
- **Lines:**
  - \( d=0 \) (blue)
  - \( d=1 \) (green)
  - \( d=2 \) (red)
  - \( d=3 \) (cyan)
  - \( d=4 \) (magenta)
  - \( d=5 \) (yellow)

**Graph 2:**
- **Axes:**
  - x-axis: \( x \)
  - y-axis: \( P(D \in S|x) \)
- **Colors:**
  - \( S=\{4,5\} \) (unsafe) (red)
  - \( S=\{2,3\} \) (unsafe*) (orange)
  - \( S=\{0,1\} \) (safe) (green)
- **Legend:**
  - **True Positives:**
    - Unknown condition
    - Predicted condition
    - Observed condition
  - **False Positives:**
    - 0%
  - **False Negatives:**
    - 0%
- **Legend Key:**
  - Unknown condition
  - Predicted condition
  - Observed condition

**Legend:**
- **Number of inspections performed [%]:**
  - 0 to 100
- **Graph Title:**
  - False +: 0% | False -: 0%
- **X-axis:**
  - Number of inspections performed
- **Y-axis:**
  - [%]
Introduction

Generative classifier

Logistic Regression

Gaussian Process Classification (GPC)

NN

Summary

Example

Professor: J-A. Goulet

Classification methods | V1.4 | Machine Learning for Civil Engineers

Polytechnique Montréal

21 / 51
Introduction

Generative classifier
Logistic Regression
Gaussian Process Classification (GPC)
NN
Summary

Example

\[ f_{X|D}(x|d) \]

\( d = 0 \) \( d = 1 \) \( d = 2 \) \( d = 3 \) \( d = 4 \) \( d = 5 \)

\[ P(D \in S|x) \]

\[ S = \{4,5\} \text{ (unsafe)} \]

\[ S = \{2,3\} \text{ (unsafe*)} \]

\[ S = \{0,1\} \text{ (safe)} \]

Number of inspections performed [%]

False +: 0% | False -: 0%

Unknown condition
Predicted condition
Observed condition

Professor: J-A. Goulet
Polytechnique Montréal

Classification methods | V1.4 | Machine Learning for Civil Engineers
**Generative classifier**

- Logistic Regression
- Gaussian Process Classification (GPC)
- NN

**Summary**

**Example**

- $f_{X|D}(x|d)$
- $P(D \in S|x)$

**False +**: 0% | False −: 0%

**Number of inspections performed**

**Unknown condition**

**Predicted condition**

**Observed condition**

---

Professor: J-A. Goulet
**Example**

- **Generative classifier**
- **Logistic Regression**
- **Gaussian Process Classification (GPC)**
- **NN**
- **Summary**

---

**Classification**

- **V1.4**
- **Machine Learning for Civil Engineers**

---

**Legend**
- **Unknown condition**
- **Predicted condition**
- **Observed condition**

---

**Example**

- **P(d ∈ S|x)**
- **f_{X|D}(x|d)**
- **False +: 0% | False -: 0%**
- **Number of inspections performed [%]**

---

**Professor:** J-A. Goulet

**Polytechnique Montréal**
**Generative classifier**

Logistic Regression

Gaussian Process Classification (GPC)

NN

**Summary**

**Example**

**Professor: J-A. Goulet**

Classification methods | V1.4 | Machine Learning for Civil Engineers

---

---
Example

Classification methods

Professor: J-A. Goulet
Polytechnique Montréal

Machine Learning for Civil Engineers

Page 21/51
**Classification methods**

*Generative classifier*
- Logistic Regression
- Gaussian Process Classification (GPC)
- NN

**Example**

- $f_{X|D}(x|d)$
- $P(D \in S|x)$

- $S = \{4, 5\}$ (unsafe)
- $S = \{2, 3\}$ (unsafe*)
- $S = \{0, 1\}$ (safe)

- False +: 0.3% | False -: 3.9%

**Number of inspections performed**

- Unknown condition
- Predicted condition
- Observed condition

---

**Professor:** J-A. Goulet

**Polytechnique Montréal**
Introduction

Generative classifier
Logistic Regression
Gaussian Process Classification (GPC)
NN
Summary

Example

- $f_{X|D}(x|d)$ for different $d$ values
- $P(D \in S|x)$ for different conditions $S$
- False +: 0.3% | False -: 3.9%
- Number of inspections performed

Professor: J-A. Goulet
Example

```
f_X|D(x|d)
```

```
P(D∈S|x)
```

```
S={4,5} (unsafe)
S={2,3} (unsafe*)
S={0,1} (safe)
```

```
False +: 0.3% | False -: 3.9%
```

```
[Unknown condition
Predicted condition
Observed condition]
```

```
Number of inspections performed
```

```
[Unknown condition
Predicted condition
Observed condition]
```

```
Professor: J-A. Goulet
Classification methods | V1.4 | Machine Learning for Civil Engineers
Polytechnique Montréal
```
Introduction

Generative classifier

Logistic Regression

Gaussian Process Classification (GPC)

NN

Summary

Example

\[ f_{X|D}(x|d) \]

\[ P(D \in S|x) \]

0 0.2 0.4 0.6 0.8 1

0 0.5 1

False +: 1.3% | False -: 3.9%

Unknown condition

Predicted condition

Observed condition

Professor: J-A. Goulet

Polytechnique Montréal

Classification methods | V1.4 | Machine Learning for Civil Engineers
### Introduction

**Generative classifier**
- Logistic Regression
- Gaussian Process Classification (GPC)
- NN

**Summary**

**Example**

**Figure:**
- Graph showing $f_{X|D}(x|d)$ for different $d$ values.
- Heatmap indicating $P(D \in S|x)$ for different $S$ sets.
- Bar chart showing the number of inspections performed and percentage of false + and false -.

**Legend:**
- Red: $S=\{4,5\}$ (unsafe)
- Orange: $S=\{2,3\}$ (unsafe*)
- Green: $S=\{0,1\}$ (safe)

False +: 1.3% | False -: 3.9%

**Professor:** J-A. Goulet

**Polytechnique Montréal**
Introduction

Generative classifier

Logistic Regression

Gaussian Process Classification (GPC)

NN

Summary

Example

Professor: J-A. Goulet

Classification methods | V1.4 | Machine Learning for Civil Engineers
Classification methods

Professor: J-A. Goulet
Polytechnique Montréal
Example

\[
f_{X|D}(x|d)
\]

\[
P(D \in S|x)
\]

False +: 1.3% | False -: 3.9%

Number of inspections performed

Unknown condition
Predicted condition
Observed condition

Professor: J-A. Goulet

Classification methods | V1.4 | Machine Learning for Civil Engineers
Example

Classification methods

- Generative classifier
- Logistic Regression
- Gaussian Process Classification (GPC)
- NN

Professor: J-A. Goulet
Polytechnique Montréal
Example

- Generative classifier
- Logistic Regression
- Gaussian Process Classification (GPC)
- NN

### Introduction

**Generative classifier**

**Logistic Regression**

**Gaussian Process Classification (GPC)**

**NN**

**Summary**

---

**Classification methods**

- V1.4
- Machine Learning for Civil Engineers

---

**False +: 1.3% | False -: 3.9%**

---

**Number of inspections performed**

- Unknown condition
- Predicted condition
- Observed condition
Introduction

Generative classifier

Logistic Regression

Gaussian Process Classification (GPC)

NN

Summary

Example

Professor: J-A. Goulet

Polytechnique Montréal

Classification methods | V1.4 | Machine Learning for Civil Engineers
### Example

**Introduction**

- Generative classifier
- Logistic Regression
- Gaussian Process Classification (GPC)
- NN

**Summary**

- Example

**Figure**: 

- Graph showing the distribution of $f_{X|D}(x|d)$ for different values of $d$.
- Heatmap indicating unsafe and safe conditions.
- Graph showing the number of inspections performed and the percentage of false positives and false negatives.

**Professor**: J-A. Goulet

**Polytechnique Montréal**

**Classification methods** | V1.4 | Machine Learning for Civil Engineers
Example

\begin{itemize}
  \item Generative classifier
  \item Logistic Regression
  \item Gaussian Process Classification (GPC)
  \item NN
\end{itemize}

\begin{itemize}
  \item Introduction
  \item Summary
\end{itemize}

\begin{figure}
\centering
\includegraphics[width=\columnwidth]{example.png}
\caption{Example Graph}
\end{figure}

\begin{itemize}
  \item Classification methods
\end{itemize}
### Introduction

#### Generative classifier

- Logistic Regression
- Gaussian Process Classification (GPC)
- NN

#### Summary

**Example**

- $f_{X|D}(x|d)$
- $P(D \in S|x)$

**False +:** 1.3% | **False −:** 3.9%

- Unknown condition
- Predicted condition
- Observed condition

---

**Professor:** J-A. Goulet

**Classification methods | V1.4 | Machine Learning for Civil Engineers**
Introduction

Generative classifier

Logistic Regression

Gaussian Process Classification (GPC)

NN

Summary

Example

Professor: J-A. Goulet

Polytechnique Montréal

Classification methods | V1.4 | Machine Learning for Civil Engineers
Introduction

Generative classifier

Logistic Regression

Gaussian Process Classification (GPC)

NN

Summary

Example

\[ f_{X|D}(x|d) \]

\[ P(D \in S|x) \]

False +: 1.3% | False -: 3.9%

Number of inspections performed

%: [0, 50, 100]

Unknown condition

Predicted condition

Observed condition

Professor: J-A. Goulet

Polytechnique Montréal
**Introduction**

*Generative classifier*  
*Logistic Regression*  
*Gaussian Process Classification (GPC)*  
*NN*  
*Summary*

---

**Example**

$$f_{X|D}(x|d)$$

<table>
<thead>
<tr>
<th>d = 0</th>
<th>d = 1</th>
<th>d = 2</th>
<th>d = 3</th>
<th>d = 4</th>
<th>d = 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.2</td>
<td>0.4</td>
<td>0.6</td>
<td>0.8</td>
<td>1</td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$$P(D \in S|x)$$

- **S = {4, 5} (unsafe)**
- **S = {2, 3} (unsafe*)**
- **S = {0, 1} (safe)**

---

**False +: 1.3% | False -: 3.9%**

---

**Number of inspections performed**

- **Unknown condition**
- **Predicted condition**
- **Observed condition**
Classification methods
**Example**

**Generative classifier**

**Logistic Regression**

**Gaussian Process Classification (GPC)**

**NN**

**Summary**

**Classification methods**

| V1.4 | Machine Learning for Civil Engineers |

---

**Professor:** J-A. Goulet

**Polytechnique Montréal**

| 21 / 51 |
J. Goulet, C. Michel, and A. Der Kiureghian

“Data-driven post-earthquake rapid structural safety assessment”
Classic Generative Methodologies

(a) Naïve Bayes

\[ f(x|y = -1) \]
\[ f(x|y = +1) \]

(b) LDA

\[ \mathcal{N}(x; \mu_{x,j}, \Sigma_{x,j}) \]

(c) QDA

\[ \mathcal{N}(x; \mu_{x,j}, \Sigma_{x,j}) \]

Method | \( f(x|y = j) \)
--- | ---
Naïve Bayes | \( X_i|y \perp \perp X_k|y, \forall i \neq k \)
| e.g. \( \mathcal{B}(x_1; \alpha, \beta) \cdot \mathcal{N}(x_2, \mu, \sigma) \)
LDA | \( \mathcal{N}(x; \mu_{x,j}, \Sigma_{x}) \)
QDA | \( \mathcal{N}(x; \mu_{x,j}, \Sigma_{x,j}) \)
Summary - Generative classifier

**MLE/MAP point estimation**

\[
p(y|x, \theta^*) = \frac{f(x|y, \theta^*) \cdot p(y; \theta^*)}{f(x; \theta^*)}
\]

**Bayesian estimation**

\[
p(y|x, D) = \int \frac{f(x|y, \theta) \cdot p(y; \theta)}{f(x; \theta)} \cdot f(\theta|D) \, d\theta
\]
Summary - Strength & Limitations

\[
p(y|x) = \frac{\text{likelihood} \cdot \text{prior}}{\text{norm. cte.}}
\]

**Strength**

- Compatible with any type of observation (error type/observation type)
- Small dataset → Bayesian inference
- Can handle unbalanced observation classes

**Limitations**

- Complex to set-up tailored model structures
- Simple structures (e.g. Naive Bayes, LDA & QDA) are outperformed by other discriminative methods
Section Outline

Logistic Regression

3.1 Revision – Linear Regression
3.2 Linear → Logistic Regression
3.3 Exemple
3.4 Limitations & summary
Revision – Linear Regression

Data
\[ \mathcal{D} = \{(x_i, y_i), \forall i = 1 : D\} \]

\[ x_i \in \mathbb{R} : \begin{cases} \text{Covariate} \\ \text{attribute} \\ \text{regressor} \end{cases} \]

\[ y_i \in \mathbb{R} : \text{Observation} \]

Model
\[ g(x) \equiv \text{fct}(x) \]

Hypothesis: \( g(x) \equiv \text{reality}, \) i.e. no variability

\[ y = g(x) + \nu, \quad \nu : \mathcal{N}(\nu; 0, \sigma^2) \]
Introduction to Logistic regression

**Data**

\[ D = \{(x_i, y_i), \forall i = 1 : D\} \]

- Covariate attribute
- regressor

\[ y_i \in \{-1, 1\} : \text{Observation} \]

\[ \Pr(\mathbf{x}) = \sigma(b_0 + b_1 x) \]

\[ = \frac{1}{1 + \exp(-z)} \]

\[ = \sigma(z) \]

\[ \sigma(z) \]

\[ 0 \]

\[ -6 \]

\[ -4 \]

\[ -2 \]

\[ 0 \]

\[ 2 \]

\[ 4 \]

\[ 6 \]
From linear to logistic regression

\[ x \in \mathbb{R}^X \rightarrow g(x) = Xb \in \mathbb{R} \rightarrow \sigma(g(x)) \in (0, 1) \]

\[ \Pr(Y = y | x) \]

\[ \sigma(z) = \frac{1}{1 + \exp(-z)} \]
Example - Linear regression & sigmoid
Likelihood

Marginal likelihoods
For \( x_i : y_i = +1 \rightarrow X \)

\[
\Pr(Y = +1|x_i) = \sigma(Xb) \equiv \sigma(x_i)
\]

For \( x_i : y_i = -1 \rightarrow X \)

\[
\Pr(Y = -1|x_i) = 1 - \sigma(Xb)
\]

Joint Log-likelihood

\[
\ln p(D|b) = \sum \ln(\sigma(X_{(+1)}b)) + \sum \ln(1 - \sigma(X_{(-1)}b))
\]

Maximum Likelihood Estimation

\[
b^* = \arg \max_b \ln p(D|b)
\]
Example - Logistic regression
Example - Logistic regression [1]

Example - Logistic regression [1]
Example - Logistic regression [1]
Limitations

▶ △ The performance depends on the capacity to **hand-pick the correct transformation functions**
   (Difficult when $D > 1 \rightarrow \text{Cross-validation})$

▶ △ No analytic formulation to estimate parameters

▶ △ Only compatible with error-free* direct observations

▶ △ Not as powerful as modern methods

Why is it used then?

▶ △ Simple

▶ △ Interpretability of model parameters $b$
   (Discrete choice models $\rightarrow$ transportation & economics)
Section Outline

Gaussian Process Classification (GPC)

4.1 GPR v.s. GPC
4.2 Updating a GP using exact observations
4.3 GPC formulation
4.4 GPML
4.5 Summary
Gaussian Process Regression summary

Given a system response so that

\[
g_i = g(x_i) \in \mathbb{R}, \quad x_i = [x_1, x_2, \ldots, x_X]^T \in \mathbb{R}^X
\]

**Data:** \( D = \{(x_i, g_i), \forall i = 1 : D\} \)

\[
g = [g_1, g_2, \ldots, g_D]^T_{1 \times D} \in \mathbb{R}^D, \quad x = [x_1, x_2, \ldots, x_X]_{D \times X}
\]

**Gaussian process:** \( g(x) : G \sim \mathcal{N}(g(x); m_G, \Sigma_G) \)

Set of discrete Gaussian random variables for which the pairwise correlation between \( G_i \) and \( G_j \) is a function of the distance between attributes

\[
[\Sigma_G]_{ij} = \rho(x_i, x_j)\sigma_{G_i}\sigma_{G_j}, \quad \rho(x_i, x_j) = \text{fct}(|x_i - x_j|)
\]
Correlation function
Updating a GP using exact observations

Given $\mathcal{D} = \{(x_i, g_i), \ i = 1, \cdots, D\}$ a set of $D$ observations and $x_*$ a set of $X_*$ covariates for which we want to predict

$$f(g_* | x_*, \mathcal{D})$$

**Reminder:** Gaussian conditionals are also gaussian

$$\begin{bmatrix} G \\ G_* \end{bmatrix}, \ m = \begin{bmatrix} m_G \\ m_{G*} \end{bmatrix}, \ \Sigma = \begin{bmatrix} \Sigma_G & \Sigma_{G*} \\ \Sigma_{G*}^T & \Sigma_* \end{bmatrix}$$

Prior knowledge

$$\begin{bmatrix} \Sigma_G \\ \Sigma_{G*} \end{bmatrix}_{ij} = \rho(x_i, x_j)\sigma_G^2$$

$$\begin{bmatrix} \Sigma_G \\ \Sigma_{G*} \end{bmatrix}_{ij} = \rho(x_i, x_*j)\sigma_G^2$$

$$\begin{bmatrix} \Sigma_* \\ \Sigma_* \end{bmatrix}_{ij} = \rho(x_*i, x_*j)\sigma_G^2$$

Posterior knowledge

$$f(g_* | x_*, \mathcal{D}) = \mathcal{N}(g_*; m_*|\mathcal{D}, \Sigma_*|\mathcal{D})$$

$$m_*|\mathcal{D} = m_{G*} + \Sigma_{G*}^T \Sigma^{-1}_G (g - m_G)$$

$$\Sigma_*|\mathcal{D} = \Sigma_* - \Sigma_{G*}^T \Sigma^{-1}_G \Sigma_{G*}$$
Example - GPR v.s. GPC [m]

\[ M_G \pm 2\sigma_f \]

\[ M_G \times g_i \]

\[ \Phi(g(x)) \]

\[ \Pr(Y = 1) \]
Example - GPR v.s. GPC [m]

\[ M_G \pm 2\sigma_f \quad \rightarrow \quad M_G \quad \times \quad g_i \quad \rightarrow \quad \Phi(g(x)) \quad \rightarrow \quad \Pr(Y = 1) \]
Example - GPR v.s. GPC

\[ \Phi(\cdot) \triangleq \text{Normal CDF} \]

\[
Pr(Y = 1|x_*, D) = \int \Phi(g_*) f(g_*|x_*, D) dg_*
\]

\[
= \Phi \left( \frac{m_{G_*|G}}{\sqrt{1 + \text{var}[G_*|G]}} \right)
\]

In a classification setup, \( g \) is not observed

Only \( y_i \in \{-1, 1\} \) is observed.

For each \( y_i \in \{-1, 1\} \) we need to infer \( g_i \)
Example - GPR v.s. GPC

\[ M_G \pm 2\sigma_f - M_G \]  
\[ \Phi(g(x)) \]  
\[ y_i \]  
\[ \Pr(Y = 1 | x) \]  
\[ M_G|D \]
Example - GPC / GPML
Gaussian Process Classification Formulation

\[
\Pr(Y = 1|x_*, \mathcal{D}) = \int \Phi(g_*) \cdot f(g_*|g, x_*, \mathcal{D}) dg_* \\
= \Phi \left( \frac{m_{G*|G}}{\sqrt{1 + \text{var}[G_*|G]}} \right)
\]

\[
f(g_*|g, x_*, \mathcal{D}) \equiv \mathcal{N}(g_*; m_{G*|G}, \Sigma_{G*|G})
\] (GPR)

\[
m_{G*|G} = \Sigma_{G*} \Sigma_{G}^{-1}(m_{G|\mathcal{D}} - m_{G} - 0)
\] (conditional mean)

\[
\Sigma_{G*|G} = \Sigma_* - \Sigma_{G*} \Sigma_{G|\mathcal{D}} \Sigma_{G*}^T
\] (conditional COV)

\[
f(g|\mathcal{D}) = \frac{p(y|g)f(g|x)}{p(y|x)}
\]

\[
\propto p(y|g) \cdot f(g|x)
\]

\[
\approx \mathcal{N}(g; m_{G|\mathcal{D}}, \Sigma_{G|\mathcal{D}})
\] (Laplace approx.)
GPC - Hidden covariate estimation, \( g | D \)

\[
p(y | g) = \prod_{i=1}^{D} p(y_i | g_i) \\
= \prod_{i=1}^{D} \Phi(y_i \cdot g_i)
\]

**Laplace approximation**

\[
m_{G|D} = \arg \max_g \ln f(g | D) \\
= \Sigma_G \nabla \ln f(y | g = m_{G|D}) \\
\text{recursive estimation} \\
\text{start: } m_{G|D} = 0 \\

\[
\Sigma_{G|D} = (\Sigma_G^{-1} - H[\ln f(y | g = m_{G|D})])^{-1}
\]
GPML package - Classification

%% Data
x_obs; % Attribute values associated with observations
y_obs; % Observed/simulated values \( y_{\text{obs}} \in \{-1,1\} \)
x_pred; % Attribute values associated with predictions

%% GPML
initial_guess_l = 10; % initial guess for \( l \)
initial_guess_s_G = 1; % initial guess for \( s_G \)
hyp.cov = \text{log}([initial_guess_l, initial_guess_s_G]);
hyp.mean = 0;

%% MLE parameter estimation
hyp = minimize(hyp, @gp, -5000, @infLaplace, @meanConst, @covSEard, @likErf, x_obs, y_obs);

%% GP Classification
[~,~,~,~,log_p] = gp(hyp, @infLaplace, @meanConst, @covSEard, @likErf, x_obs, y_obs, x_pred, ones(n, 1));

%% Estimated values
Pr_y = \text{exp}(log_p); % \text{Pr}(Y=1|x^*,D)
Example - GPC / GPML
Example - GPC / GPML
Example - GPC / GPML
Summary - Gaussian Process Classification

**Strength**

- Simple to set-up for any number of covariates \( x \)
- Allows interpolating and extrapolating (uncertainty estimates)

**Limitations**

- Only compatible with error-free direct observations
- Whenever possible choose *regression* over *classification*, e.g. Soil contamination example
- The performance is limited for large datasets \( D > 10^5 \)
Section Outline

NN

5.1 Introduction
5.2 Structure
5.3 Deep Learning
5.4 Summary
Introduction to the Neural Network Structure

**Data** \( D = \{ (x_i, y_i), \forall i = 1 : D \} \)

\( x_i \in \mathbb{R} : \)  
\[ \begin{cases} \text{Covariate} \\ \text{attribute} \\ \text{regressor} \end{cases} \]

\( y_i \in \{-1, 1\} : \) Observation

\[
Pr(Y = 1|x) = \sigma(w_0 + w_1x_1 + w_2x_2 + w_3x_3)
\]

\[
= \frac{1}{1 + \exp(-z)}
\]

\[
\sigma(z)
\]

---

**Professor:** J-A. Goulet
Neural Network Structure
Neural Network Structure
Neural Network Structure
Neural Network Structure
Deep learning refers to a neural network which has a deep structure

i.e., a large number of hidden layers
Summary - Neural Networks

**Strength**

- Deep Learning is the *state-of-the-art* method
- Unequaled regression/classification accuracy

**Limitations**

- In order to outperform other methods (e.g. GPR/GPC) it requires many hidden activation units
  - large number of parameters to learn, > $10^6$
- Learning a large number of parameters
  - requires a large amount of data $\gg 10^6$
Summary

**Generative Classifier:**
- Offers the most flexibility (e.g. indirect observations)
- Best suited for small datasets

**Gaussian Process Classification:**
- Quick & easy
- Allows interpolating and extrapolating (uncertainty estimates)
- Best suited for medium size datasets

**Neural networks / Deep Learning:**
- State-of-the-art method
- Best suited for large datasets