Machine Learning
EDS
Tutorial: Week 2

Janneke van Brummelen

Vrije Universiteit Amsterdam

Period 3.4
2023/2024

This tutorial

1. Short recap of this week's material

2. A small quiz

3. Discuss Problems 1, 2.1 and 3 of Problem set 2

This week:

- We **formalized** the learning problem:
  - We imposed that $D_n = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$ with $(X_i, Y_i) \sim P$ and independent over i.
  - Predictors $f : \mathcal{X} \to \mathcal{Y}$
  - Cost functions $c : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$
  - Generalization error/risk: $\mathcal{R}^c_P(f) = \mathbb{E}[c(f(X), Y)|D_n]$

- We discussed **ideal prediction**:
  - Bayes risk: $\mathcal{R}^*_P = \inf_{f \in \mathcal{F}} \mathcal{R}^c_P(f)$
  - Bayes predictors/target functions: any $f^* \in \mathcal{F}$ for which $\mathcal{R}^c_P(f^*) = \mathcal{R}^*_P$
  - Excess risk: $\ell(f^*, f) = \mathcal{R}^c_P(f) - \mathcal{R}^*_P$
  - We derived Bayes risk and Bayes predictors for two fundamental frameworks:
    - Regression with quadratic cost
    - Classification with 0-1 loss

Table of Contents

1 Short quiz

2 Problem 1

3 Problem 2.1

4 Problem 3

Quiz: question 1

Say you are trying to predict a variable of interest based on 10 different categorical variables. What kind of supervised machine learning problem are you facing?

A Classification problem

B Regression problem

C **We cannot say based on the provided information**

Quiz: question 2

Is the following statement true or false?

"*A predictor $f \in \mathcal{F}$ with strictly positive excess risk, $\ell(f^*, f) > 0$, cannot be a Bayes predictor*"

- **True**
- False

Quiz: question 3

Is the following statement true or false?

"If there is a deterministic relationship between X and Y, i.e. if there exists a function $g : \mathcal{X} \to \mathcal{Y}$ such that $g(X) = Y$ a.s., then Bayes risk is equal to zero."

- **True**
- False

Table of Contents

1 Short quiz

2 Problem 1

3 Problem 2.1

4 Problem 3

Problem 1: Die roll experiment

Consider two 6-face dice, die 1 and die 2:

- Die 1: 3 blue faces, 3 red faces
- Die 2: 5 blue faces, 1 red face

First choose a die at random, say die 1 with probability $p \in [0, 1]$, and then roll the selected die. The color appearing on the upward face is recorded.

Denote by $X \in \{1, 2\}$ the selected die, and $Y \in \{0, 1\}$ the outcome of rolling the die, say $Y = 1$ for blue, $Y = 0$ for red.

Let us study the problem of predicting the outcome Y given X.

Problem 1: Die roll experiment

1.1 Is this a regression or a classification problem? Why?

1.2 Show that $\eta(X) = \mathbb{P}(Y = 1|X)$, where $\eta(X) = \mathbb{E}[Y|X]$.

1.3 For the 0-1 cost function, what is the general form (expressed in terms of η) of Bayes predictors and Bayes risk in this framework?

1.4 Derive η and Bayes predictors explicitly, and then show that Bayes risk is equal to $\mathcal{R}_P^* = \dfrac{1}{6} + \dfrac{p}{3}$.

1.5 Intuitively, why is Bayes risk an increasing function of p?

Table of Contents

1 Short quiz

2 Problem 1

3 Problem 2.1

4 Problem 3

Problem 2

Consider the regression framework with $\mathcal{Y} = \mathbb{R}$. Let $F_{Y|X}(\cdot|x)$ be the conditional distribution function of $Y|X = x$. In this exercise, we assume the existence of the conditional PDF $f_{Y|X=x}(\cdot|x)$.

Problem 2.1

2.1 For the absolute value cost $c(y, y') = |y - y'|$ for $y, y' \in \mathcal{Y}$, let us show that $f^* : x \mapsto \text{med}(Y|X = x)$ is a Bayes predictor, where $\text{med}(Y|X = x)$ stands for the median of the distribution of Y given $X = x$:

(a) Fix $x \in \mathcal{X}$, and consider the function $r_x : a \mapsto r_x(a) = \mathbb{E}\left[|a - Y|\middle|X = x\right]$. Show that for any $a \in \mathbb{R}$, $r_x'(a) = 2\mathbb{P}(Y \leq a|X = x) - 1$ and $r_x''(a) = 2f_{Y|X=x}(a)$.

(b) Show that r_x admits a global minimum at $a_x^* = \text{med}(Y|X = x)$.

(c) Deduce that $\mathbb{E}\left[|f^*(X) - Y|\right] \leq \mathbb{E}\left[|f(X) - Y|\right]$ for any predictor $f \in \mathcal{F}$ and conclude that $f^* : x \mapsto \text{med}(Y|X = x)$ is a Bayes predictor.

Hint for 2.1(a)

Hint: Recall that for $\varphi(x) = \int_{A(x)}^{B(x)} g(x,t)dt$ *(assuming all functions are differentiable),*

$$\varphi'(x) = \int_{A(x)}^{B(x)} \frac{\partial g(x,t)}{\partial x} dt + B'(x) g\big(x, B(x)\big) - A'(x) g\big(x, A(x)\big)$$

This rule cannot be applied immediately to improper integrals. To be able to apply the rule, you could split the integrals

$$\int_a^\infty y f_{Y|X}(y|x)\ dy = \int_a^M y f_{Y|X}(y|x)\ dy + \int_M^\infty y f_{Y|X}(y|x)\ dy.$$

Problem 2.2

2.2 Derive a Bayes predictor when c is the asymmetric absolute value cost:
$c(y, y') = c_-(y' - y)\mathbb{1}_{y<y'} + c_+(y - y')\mathbb{1}_{y \geq y'}$ for $y, y' \in \mathcal{Y}$,
with $c_-, c_+ \geq 0$, $c_- + c_+ > 0$.
How does it relate to the Bayes predictor of the previous question?

I will skip the derivations because almost the same as 2.1, but do try it yourself!

Problem 2.2: final answer

When using the same approach as for Problem 2.1, you will find that the following predictor is a Bayes predictor:

$$f^* : x \mapsto F^{-1}_{Y|X=x}(q^*) \quad \text{with} \quad q^* := \frac{c_-}{c_- + c_+},$$

where $F^{-1}_{Y|X=x}(q)$ denotes the q-quantile of the distribution of Y conditional on $X = x$.

In other words, for any $q \in [0, 1]$:

$$F^{-1}_{Y|X=x}(q) = a \quad \iff \quad F_{Y|X=x}(a) = \mathbb{P}(Y \leq a|X = x) = q$$

(or equivalently if $\mathbb{P}(Y > a|X = x) = 1 - q$)

Table of Contents

1 Short quiz

2 Problem 1

3 Problem 2.1

4 Problem 3

Problem 3

Let $\mu_0, \mu_1 \in \mathbb{R}^d$ two vectors and Σ_0, Σ_1 two symmetric positive definite matrices of size d. Let Y be a random variable such that $\mathbb{P}(Y = 1) = 1 - \mathbb{P}(Y = 0) = p \in [0, 1]$, and let X be a random variable such that $X|Y = i \sim \mathcal{N}(\mu_i, \Sigma_i)$, for $i \in \{0, 1\}$. We consider the classification problem of predicting Y given X with the 0-1 cost.

1. Recall the general form of Bayes classifiers in this framework (expressed in terms of η).
2. Show that $\eta(x) > 1/2 \iff \dfrac{\mathbb{P}(Y = 1|X = x)}{\mathbb{P}(Y = 0|X = x)} > 1$.
3. For $i = 0, 1$, derive an expression of $\mathbb{P}(Y = i|X = x)$ in terms of $N_i(x)$ (the pdf of a multivariate normal with mean μ_i and variance Σ_i), $f_X(x)$ and p.

Problem 3

4. Show that a Bayes classifier can be expressed as

$$f^* : x \in \mathbb{R}^d \mapsto \mathbb{1}_{x^\top A x + b^\top x + c > 0},$$

for some matrix A, vector b and constant c, depending on $\mu_0, \mu_1, \Sigma_0, \Sigma_1$ and p.
Hint: Use the pdf of the multivariate normal distribution.

5. What is the geometric form of the decision boundary for this classifier? What if $\Sigma_0 = \Sigma_1$?

Problem 3(e)

The decision boundary $x^\top A x + b^\top x + c = 0$ is the equation of a quadric surface.

For instance, for $d = 2$, it can take the following forms (conic sections):