Machine Learning
EDS
Tutorial: Week 1

Janneke van Brummelen

Vrije Universiteit Amsterdam

Period 3.4
2023/2024

Practicalities

- You can find problems that will be discussed during tutorial on Canvas.
- Solutions will be posted after each tutorial.
- Advisable to come to tutorials, because will get more elaborate answers here.
- Strong advice: make the exercises before tutorial.
- Before we dive into exercises: warm-up quiz.

Table of Contents

1 Warm-up Quiz

2 Problem 1

3 Problem 2

4 Problem 3

Quiz: question 1

Consider the function $f : x \in \mathbb{R} \mapsto |x|$.

Which of the following statements is true?

A The function f is convex and concave.

B The function f is strictly convex.

C **The function f is convex, but not strictly convex and not concave.**

Quiz: question 2

Consider the function $f : x \in \mathbb{R}_{++} \mapsto \ln(x)$, where $\mathbb{R}_{++} = (0, \infty)$.

Which of the following statements is true?

- A The function f is convex, but not strictly convex.
- B The function f is concave, but not strictly concave.
- C The function f is strictly convex.
- D **The function f is strictly concave.**

Quiz: question 3

Consider some function $f : S \to \mathbb{R}$ defined on the closed interval $S = [-10, 10]$. Is the following statement is true or false?

"The point $x^* \in S$ can only be a global minimum of f if it is a critical point."

- True
- **False**

Quiz: question 4

Consider some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Consider two events $A, B \in \mathcal{F}$ for which $A \cap B = C$ and $\mathbb{P}(C) > 0$.

Which of the following statements must be true?

A $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$

B $\mathbb{P}(A \cup B) < \mathbb{P}(A) + \mathbb{P}(B)$

C $\mathbb{P}(A \cap B) < \mathbb{P}(A) + \mathbb{P}(B)$

(Both B and C are correct)

Table of Contents

1 Warm-up Quiz

2 Problem 1

3 Problem 2

4 Problem 3

Problem 1.1

(Markov inequality) Let X be a real-valued non-negative random variable. Show that:

$$\forall a > 0, \quad \mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}.$$

Problem 1.2

(Corollary for sub-Gaussian random variables) A real-valued random variable X is said to be b-sub-Gaussian, $b > 0$, if for any $s \in \mathbb{R}$ it holds that: $\mathbb{E}[\exp(sX)] \leq \exp\left(\frac{s^2 b^2}{2}\right)$. It can be shown that any b-sub-Gaussian random variable X has $\mathbb{E}[X] = 0$ and $\mathbb{V}\mathrm{ar}(X) \leq b^2$. Assume that X is b-sub-Gaussian and let $a > 0$.

(a) First show that

$$\forall s > 0, \quad \mathbb{P}\big(X \geq a\big) \leq e^{\frac{s^2 b^2}{2} - sa}.$$

(b) Deduce that

$$\mathbb{P}(X \geq a) \leq \exp\left(-\frac{a^2}{2b^2}\right).$$

Hint: Notice that left-hand side in (a) does not depend on s.

Table of Contents

1 Warm-up Quiz

2 Problem 1

3 Problem 2

4 Problem 3

Problem 2.1

(Sum of independent random variables) Let $X_1, \ldots, X_n$ be n real-valued random variables. Assume that for any $i \leq n$, X_i is b_i-sub-Gaussian for some positive constant b_i.
Let $S_n = X_1 + \ldots + X_n$.

(a) Show that S_n is b-sub-Gaussian for some explicit positive constant b expressed in terms of $b_1, \ldots, b_n$.

(b) Deduce the concentration inequalities

$$\forall a > 0, \quad \mathbb{P}(S_n \geq a) \leq \exp\left(-\frac{a^2}{2\sum_{i=1}^n b_i^2}\right), \quad \text{and} \quad \mathbb{P}(S_n \leq -a) \leq \exp\left(-\frac{a^2}{2\sum_{i=1}^n b_i^2}\right).$$

(c) Deduce a concentration inequality for $|S_n|$.
Hint: Notice that $\{|S_n| \geq a\}$ *and* $\{S_n \geq a\} \cup \{S_n \leq -a\}$ *are the same event.*

Problem 2.2

(Gaussian case)

(a) Let X be a Gaussian random variable $X \sim \mathcal{N}(\mu, \sigma^2)$. Show that $X - \mathbb{E}[X]$ is b-sub-Gaussian for some explicit positive constant b expressed in terms of σ.

(b) Deduce a concentration inequality for a sum of n independent Gaussian random variables $S_n = \sum_{i=1}^{n} \Big(X_i - \mathbb{E}[X_i] \Big)$, where $X_i \overset{i.i.d.}{\sim} \mathcal{N}(\mu_i, \sigma_i^2)$.

Table of Contents

1 Warm-up Quiz

2 Problem 1

3 Problem 2

4 Problem 3

Problem 3 (Hoeffding's inequality)

In this exercise, we will derive **Hoeffding's inequality**:
provides an upper bound on the probability that the sum of bounded independent random variables deviate from its expectation by more than some amount.

Say we have n independent random variables $X_i \in [c_i, d_i]$, and we denote their sum by $S_n = X_1 + \cdots + X_n$, then Hoeffding's inequality reads for any $t > 0$:

$$\mathbb{P}(S_n - \mathbb{E}(S_n) \geq t) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^{n}(d_i - c_i)^2}\right) ,$$

Problem 3 (Hoeffding's inequality)

Let X be a bounded random variable such that $X \in [c, d]$, for some constants $c < d$, and $\mathbb{E}[X] = 0$. The goal is to show that X is b-sub-Gaussian for some explicit parameter b, and deduce concentration inequalities for sum of independent bounded random variables.

3.1 First show that for any $x \in [c, d]$ and $t \in \mathbb{R}$,

$$e^{tx} \leq \frac{d-x}{d-c} e^{tc} + \frac{x-c}{d-c} e^{td}$$

Hint: Use $x = \frac{d}{d-c}x - \frac{c}{d-c}x = \frac{d-x}{d-c}c + \frac{x-c}{d-c}d$ and convexity of $f(x) = e^x$.

3.2 Deduce that

$$\mathbb{E}[e^{tX}] \leq \frac{d}{d-c} e^{tc} + \frac{-c}{d-c} e^{td}.$$

Problem 3 (Hoeffding's inequality)

3.3 Letting $h := t(d-c)$, $p := \dfrac{-c}{d-c}$ and $L : h \mapsto L(h) := -hp + \ln(1 - p + pe^h)$, check that

$$e^{L(h)} = \frac{d}{d-c} e^{tc} + \frac{-c}{d-c} e^{td}.$$

which implies that $\mathbb{E}[e^{tX}] \leq e^{L(h)}$.

3.4 Considering the function $L : h \mapsto L(h)$, show that $L(0) = L'(0) = 0$, and that for any h, $L''(h) \leq 1/4$.

3.5 Using a second order Taylor approximation, deduce that for any h:

$$L(h) \leq \frac{h^2}{8}.$$

Problem 3 (Hoeffding's inequality)

3.6 Conclude that X is b-sub-Gaussian for some explicit positive constant b expressed in terms of c and d.

3.7 How is the previous result modified if $\mathbb{E}[X] \neq 0$?

3.8 Letting $X_1, \ldots, X_n$ be n independent bounded random variables such that for any i, $X_i \in [c_i, d_i]$, deduce concentration inequalities for $S_n = \sum_{i=1}^{n}(X_i - \mathbb{E}[X_i])$ and $|S_n|$.