Theory of Machine Learning

Exercise sheet 7 — Session 7

Exercise I (On some inequalities). The objective of this exercise is to prove some classical bounds of probability theory.

1. (Markov's inequality) Given a non-negative random variable X, show that:

$$\forall t > 0, \qquad \mathbb{P}\left(X \geq t\right) \leq \frac{\mathbb{E}\left[X\right]}{t}\,.$$

(Hint: any real number x can be represented as $x = x\mathbb{1}_{x<t} + x\mathbb{1}_{x\geq t}$.)

2. (Chebyshev's inequality) Given an integrable random variable X with expected value μ and variance σ^2, show that:

$$\forall t > 0, \qquad \mathbb{P}\left(|X-\mu| \geq t\right) \leq \frac{\sigma^2}{t^2}\,.$$

(Hint: apply Markov's inequality to the random variable $(X-\mu)^2$.)

3. (Generic Chernoff bound) Given a random variable X, show that:

   (a) Given $t > 0$ and $a \in \mathbb{R}$:

$$\mathbb{P}\left(X \geq a\right) = \mathbb{P}\left(\exp\left(tX\right) \geq \exp\left(ta\right)\right)\,.$$

   (b) Given $a \in \mathbb{R}$:

$$\mathbb{P}\left(X \geq a\right) \leq \inf_{t>0} \mathbb{E}\left[\exp\left(tX\right)\right]\exp\left(-ta\right)$$

   (Hint: apply Markov's inequality)

Exercise II (On the Gaussian tails). In this exercise, we want to compute some bounds on the Gaussian tails of a random variable $X \sim \mathcal{N}\left(0,1\right)$.

1. Chernoff bound on the Gaussian tail:

   (a) Given $t > 0$, show that $\mathbb{E}\left[\exp\left(tX\right)\right] = \exp\left(\frac{t^2}{2}\right)$.

   (b) Using Question 3.b. of Exercise II, show that $\mathbb{P}\left(X \geq a\right) \leq \inf_{t>0}\exp\left(\frac{t^2}{2} - ta\right)$ with $a \in \mathbb{R}$.

   (c) Minimize in $t > 0$, the following polynomial $t \mapsto \frac{t^2}{2} - ta$ with $a \in \mathbb{R}$.

   (d) Show that $\mathbb{P}\left(X \geq a\right) \leq \exp\left(-a^2/2\right)$ with $a \in \mathbb{R}$.

2. Better bound on the Gaussian tail:

   (a) Given $a > 0$, show that:

$$\frac{1}{\sqrt{2\pi}}\int_a^\infty \exp\left(-x^2/2\right)\,dx = \int_0^\infty \exp\left(-a^2/2\right)\exp\left(-ay\right)\exp\left(-y^2/2\right)\,dy\,.$$

   (Hint: make the change of variable $x = a + y$.)

   (b) Given $a > 0$, show that:

$$\int_0^\infty \exp\left(-a^2/2\right)\exp\left(-ay\right)\exp\left(-y^2/2\right)\,dy \leq \frac{1}{\sqrt{2\pi}}\exp\left(-a^2/2\right)\int_0^\infty \exp\left(-ay\right)\,dy\,.$$

   (c) Given $a > 0$, show that:

$$\int_0^\infty \exp\left(-ay\right)\,dy = \frac{1}{a}\,.$$

   (d) Finally, given that $\mathbb{P}\left(X \geq a\right) = \frac{1}{\sqrt{2\pi}}\int_a^\infty \exp\left(-x^2/2\right)\,dx$, show that:

$$\forall a \geq 1, \qquad \mathbb{P}\left(X \geq a\right) \leq \frac{1}{\sqrt{2\pi}}\exp\left(-a^2/2\right)\,.$$

Exercise III (On Gaussian vectors) ✎. Let X be a Gaussian random variable $\mathcal{N}(0,1)$ and Z be a uniformly distributed random variable on $\{-1,1\}$ independent of X.

1. Show that ZX is a Gaussian random variable.

   (Hint: A random variable X is gaussian if for every continuous and bounded function $h : \mathbb{R} \to \mathbb{R}$, we have: $\mathbb{E}[h(X)] = \int_{\mathbb{R}} h(x)f(x)\,dx$ with f the gaussian density.)

2. Show that the vector (X, ZX) is not a Gaussian vector.

   (Hint: A random vector $(X_1, \ldots, X_d) \in \mathbb{R}^d$ is said to be a Gaussian vector if any linear combination of its components is a Gaussian random variable, i.e.: for all $\alpha_1, \ldots, \alpha_d \in \mathbb{R}$, $\alpha_1 X_1 + \cdots \alpha_d X_d$ is a gaussian random variable.)

3. Compute the covariance $\mathrm{Cov}(X, XZ)$.

4. Sample $n = 1000$ Gaussian vector $(X_1, X_2) \sim \mathcal{N}(0, \mathbf{I}_2)$ and plot it in 2D.

5. Sample $n = 1000$ random vector (X, ZX) and plot it in 2D. Compare it to the previous plot. What do you observe?

Exercise IV (Expected empirical risk) ✎. Assume that $Y = \Phi\theta^\star + \varepsilon$ where ε is centered and the ε_is are independent, and have common variance σ^2 (assumptions I and II in the lecture).

1. Show that

   $$\widehat{R}(\hat{\theta}) = \frac{1}{n} \|\Pi\varepsilon\|^2 ,$$

   where $\Pi := \mathbf{I} - \Phi(\Phi^\top \Phi)^{-1}\Phi^\top \in \mathbb{R}^{n\times n}$.

2. Show that

   $$\mathbb{E}\left[\widehat{R}(\hat{\theta})\right] = \frac{n-d}{n}\sigma^2 .$$

   Hint: $\Pi := \mathbf{I} - \Phi(\Phi^\top \Phi)^{-1}\Phi^\top \in \mathbb{R}^{n\times n}$ *is an orthogonal projection matrix.*