Theory of Machine Learning

Exercise sheet 8 — July 5, 2024

Exercise I (On Popoviciu's inequality) The objective of this exercise is to prove Popoviciu's bound on the variance of an almost surely bounded random variable.

Given a square integrable random variable X such that $\mathbb{P}(m \leq X \leq M) = 1$ with $M, m \in \mathbb{R}$, Popoviciu's inequality states:

$$\mathrm{Var}(X) \leq \frac{(M-m)^2}{4}.$$

1. Proof of the inequality using Bhatia–Davis inequality:
   (a) Show that $\mathbb{E}[(M-X)(X-m)] \geq 0$.
   (b) Show that $\mathbb{E}[(M-X)(X-m)] = -\mathbb{E}\left[X^2\right] - mM + (m+M)\mathbb{E}[X]$.
   (c) Deduce that $\mathbb{E}\left[X^2\right] \leq -mM + (M+m)\mathbb{E}[X]$.
   (d) (Bhatia–Davis inequality) Knowing that $\mathrm{Var}(X) = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2$, show that:
   $$\mathrm{Var}(X) \leq (M - \mathbb{E}[X])(\mathbb{E}[X] - m).$$
   (e) (AM–GM inequality) Given $a, b \in \mathbb{R}^2$, prove that $ab \leq \left(\frac{a+b}{2}\right)^2$.
   (f) Apply the AM–GM inequality on the variance upper bound given by Bhatia–Davis inequality to get Popoviciu's inequality.

2. (The equality case) Given a discrete random variable X with $\mathbb{P}(X = m) = 1/2$ and $\mathbb{P}(X = M) = 1/2$, show that:
$$\mathrm{Var}(X) = \frac{(M-m)^2}{4}.$$

3. Application of the Popoviciu's inequality to a continuous uniformly distributed random variable $X \sim \mathcal{U}([a,b])$ with $a, b \in \mathbb{R}^2$ (and $a < b$). The density of X is:
$$\forall x \in \mathbb{R}, \qquad f(x) := \frac{1}{b-a}\mathbb{1}_{x\in[a,b]}.$$
   (a) Compute the expected value $\mathbb{E}[X] = \int_{\mathbb{R}} x f(x)\, dx$.
   (b) Compute the second moment $\mathbb{E}\left[X^2\right] = \int_{\mathbb{R}} x^2 f(x)\, dx$.
   (c) Knowing that $\mathrm{Var}(X) = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2$, compute the variance of X.
   (*Hint:* $\mathrm{Var}(X) = \frac{(b-a)^2}{12}$.)
   (d) Apply Popoviciu's inequality to bound the variance of X. What are your thoughts on the tightness of this bound?

4. Compare this bound to the one seen in the lecture when the loss function is bounded. To remind you:
$$\mathrm{Var}(\ell(Y, f(X))) \leq 4\ell_\infty^2,$$
where the loss $\ell(\cdot,\cdot)$ is bounded by $\ell_\infty \in \mathbb{R}_+$ and (X, Y) is a data sample.

Exercise II (On the generalization bound of linear regression in fixed design) The objective of this exercise is to prove a generalization bound for an OLS estimator in the fixed design setting.

Consider vector-valued inputs and real-valued outputs ($\mathcal{X} = \mathbb{R}^d$ and $\mathcal{Y} = \mathbb{R}$) with $X := (X_1, \ldots, X_n)^\top \in \mathbb{R}^{n\times d}$ the input vector and $Y := (Y_1, \ldots, Y_n)^\top \in \mathbb{R}^n$ the response vector. Let $\phi(x) = (x_1, \ldots, x_d)^\top$ and $\Phi \in \mathbb{R}^{n\times d}$ the matrix of inputs with row i defined as $\Phi_{i,:} := \phi(X_i)^\top$. We work in the fixed design setting where for a fixed input $X \in \mathbb{R}^{n\times d}$, the output is $Y = \Phi\theta + \varepsilon$ (ε i.i.d. $\mathcal{N}(0, 1)$) and $\theta \in \mathbb{R}^d$. The true risk $\mathcal{R}(\theta)$ of θ and its empirical counterpart $\widehat{\mathcal{R}}(\theta)$ are defined as:

$$\mathcal{R}(\theta) := \mathbb{E}_\varepsilon\left[\frac{1}{n}\|Y - \Phi\theta\|_2^2\right],$$
$$\widehat{\mathcal{R}}(\theta) := \frac{1}{m}\sum_{i=1}^{m}\frac{1}{n}\left\|Y^{(i)} - \Phi\theta\right\|_2^2,$$

where $\{Y^{(i)}\}_{i=1}^m$ are i.i.d. random vectors sampled as $Y^{(i)} = \Phi\theta + \varepsilon^{(i)}$ ($\varepsilon^{(i)}$ i.i.d. $\mathcal{N}(0, \mathbf{I}_n)$).

Finally, a remainder on chi-squared distribution: if $Z_1, \ldots, Z_n$ are independent standard normal $\mathcal{N}(0, 1)$ random variables, then the sum of their squares, $K := \sum_{i=1}^n Z_i^2$, is distributed according to the chi-squared distribution with n degrees of freedom, which is denoted as $K \sim \chi^2(n)$. Also, $\mathbb{E}[K] = n$ and $\operatorname{Var}(K) = 2n$.

1. Generalization bound for a fixed θ using Chebyshev's inequality:
   (a) Show that $Y_j^{(i)} \sim \mathcal{N}(\Phi_{j,:}\theta, 1)$ with $i \in [\![m]\!]$ and $j \in [\![n]\!]$.
   (b) Show that $\left\|Y^{(i)} - \Phi\theta\right\|_2^2 = \left(Y^{(i)} - \Phi\theta\right)^\top\left(Y^{(i)} - \Phi\theta\right)$ is distributed as $\chi^2(n)$.
   (Hint: rewrite the previous quantity as a sum.)
   (c) Compute the variance of $\left\|Y^{(i)} - \Phi\theta\right\|_2^2$.
   (d) Compute the variance of $\widehat{\mathcal{R}}(\theta)$.
   (Hint: the elements of the sum $\widehat{\mathcal{R}}(\theta)$ are independent.)
   (e) Using Chebyshev's inequality with $t > 0$, show that:
   $$\mathbb{P}\left(\left|\widehat{\mathcal{R}}(\theta) - \mathcal{R}(\theta)\right| > t\right) \leq \frac{\operatorname{Var}\left(\widehat{\mathcal{R}}(\theta)\right)}{t^2} = \frac{2}{mnt^2}.$$
   (f) Solve in $t > 0$, $\delta = \frac{2}{mnt^2}$.
   (g) Rewrite the previous bound as follows:
   $$\mathbb{P}\left(\left|\widehat{\mathcal{R}}(\theta) - \mathcal{R}(\theta)\right| \leq \sqrt{\frac{2}{mn\delta}}\right) \geq 1 - \delta.$$
   (h) (Numerical application) set $\delta = 0.01$ in the previous bound and rewrite it as:
   $$\mathbb{P}\left(\left|\widehat{\mathcal{R}}(\theta) - \mathcal{R}(\theta)\right| \leq 10\sqrt{\frac{2}{mn}}\right) \geq 0.99.$$