Theory of Machine Learning

Exercise sheet 4 — Session 4

Exercise I (A bit of coding) 💻. The goal of this exercise is to reproduce the figure of slide 79. 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 $Y = \Phi\theta^\star + \varepsilon$ (ε i.i.d. $\mathcal{N}\left(0, \sigma^2\right)$) and $\theta^\star \in \mathbb{R}^d$.

1. Generate the data for fixed $d = 2$, $\theta^\star = (1, \ldots, 1)^\top \in \mathbb{R}^d$ and noise $\sigma = 1$:
   (a) Sample the input data $X_{i,j} \sim \mathcal{U}([-1, 1])$, where $X \in \mathbb{R}^{n\times d}$. (*Hint: use* `numpy.random.uniform()`)
   (b) Compute the design matrix $\Phi \in \mathbb{R}^{n\times d}$.
   (c) Compute the output $Y = \Phi\theta^\star + \varepsilon$ on the fixed input data X. (*Hint: use* `numpy.random.normal()`)
2. Estimation of the expected excess risk $\mathbb{E}\left[\mathcal{R}(\hat\theta)\right] - \mathcal{R}^\star := \mathbb{E}_Y\left[\left\|\hat\theta - \theta^\star\right\|^2_{\widehat\Sigma}\right]$, where $\left\|\hat\theta - \theta^\star\right\|^2_{\widehat\Sigma} := (\hat\theta - \theta^\star)^\top \widehat\Sigma(\hat\theta - \theta^\star)$ and $\widehat\Sigma := \frac{1}{n}\Phi^\top\Phi$:
   (a) For the fixed input X from Question 1.a, generate $N = 100$ samples of Y as described in Question 1.c.
   (b) Compute the OLS estimators $\hat\theta_i := (\Phi^\top\Phi)^{-1}\Phi^\top Y$ for each sampled Y_i ($i \in [\![N]\!]$).
   (c) Compute the estimate of $\mathbb{E}\left[\mathcal{R}(\hat\theta)\right] - \mathcal{R}^\star$ as $\frac{1}{N}\sum_{i=1}^N \left\|\hat\theta_i - \theta^\star\right\|^2_{\widehat\Sigma}$.
3. By reusing the previous code that computes the (estimated) excess risk for a fixed n, plot the (estimated) excess risk as a function of $n \in [\![10, 100]\!]$.
4. (Bonus) For each n, repeat the experiment several times and plot error bars.

Exercise II (Mahalanobis distance) ✎. Let $A \in \mathbb{R}^{d\times d}$ be a positive definite matrix and set $\|u\|_A^2 := u^\top A u$ for all $u \in \mathbb{R}^d$. As in the lecture, define $d_A(x, y) := \|x - y\|_A$. Let us prove that d_A is indeed a distance.

1. Write $u^\top A u$ when $d = 2$ as a function of the coefficients of u and A.
2. Prove that d_A is symmetric;
3. Prove that $d_A(x, y)$ is always greater than 0;
4. Prove that $d_a(x, y) = 0$ only if $x = y$;
5. Prove that d_A satisfies the triangle inequality
$$\forall x, y, z \in \mathbb{R}^d, \qquad d_A(x, y) \le d_A(x, z) + d_A(z, y)\,.$$
   Hint: prove that $(x^\top A y)^2 \le x^\top A x \cdot y^\top A y$.
6. Is d_A a distance when A is only assumed to be positive *semi*-definite?

Exercise III (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.*

Exercise IV (On the Moore–Penrose inverse) ✎. The goal of this exercise is to explore fundamental properties of the Moore–Penrose inverse, as defined via the Singular Value Decomposition (SVD) in the lecture slides.

1. Given $A := \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$

   (a) Compute the Moore–Penrose inverse of A.

   (b) As seen in the lecture slide 68, compute $A^\dagger A$ and $\mathbf{I}_2 - A^\dagger A$.

   (c) What are the properties of the previous matrices?

2. Show the following equalities for $M \in \mathbb{R}^{m \times n}$:

   (a) $MM^\dagger M = M$.

   (b) $M^\dagger M M^\dagger = M^\dagger$.