Mathematical Underpinnings Lab 1 28.02.2024

Task 1 (logistic regression - generative vs discriminative approach)

a) **Generative approach**

- Generate a sample from $\mathcal{N}(m_1, \Sigma)$ $(Y = 1)$ and $\mathcal{N}(m_2, \Sigma)$ $(Y = -1)$[1]. Does the distribution of $P(Y = 1|X = x)$ correspond to a logistic model?
- If the distribution of $P(Y = 1|X = x)$ corresponds to a logistic model, find the formulas for the parameters of the logistic model (the coefficients and the intercept). Compare obtained parameters β with the estimated $\hat{\beta}$.

Simulations details: each sample from a normal distribution consists of $n = 500$ observations (in total we have 1000 observations), $m_1 = (1, 1)$, $m_2 = (0, 0)$, and

$$\Sigma = \begin{pmatrix} 1 & -0.5 \\ -0.5 & 1 \end{pmatrix}.$$

Assume that $\pi = P(Y = 1) = 0.5$.

b) **Discriminative approach**

- Generate a sample from a gaussian mixture

$$\frac{1}{2}\mathcal{N}(m_1, \Sigma) + \frac{1}{2}\mathcal{N}(m_2, \Sigma).$$

  Next, generate y_i for $i = 1, 2, \ldots, n$ from Bernoulli distribution with probability of success $1/(1 + \exp(-\beta_0 - \beta x_i)$, where $\beta_0 = -2$ and $\beta = (2, 2)$. Fit a logistic model and compare β with the estimated $\hat{\beta}$.
- Draw a scatterplot of data generated in a) and in b). Compare the distributions. What distinguishes the generative approach from the discriminative approach?

c) Generate n_1 observations from the distribution $\mathrm{Exp}(1)$[2] (corresponding to $Y = 1$) and n_2 observations from a certain distribution such that $P(Y = 1|X = x)$ corresponds to a logistic model (you may fix $n_1 = 1000$, $n_2 = 2000$ and $\beta_1 = 1$). Fit the model.
Q1. What is the distribution of $f_{X|Y=-1}$?
Q2. Once π is fixed, do we have freedom to choose an intercept in the logistic model thus constructed?
Q3. If π is not given, is the intercept in the logistic model uniquely determined?
If that is doable, given β_1 and π compute β_0.

Task 2 (population risk, empirical risk, excess risk)

We have an i.i.d. sample $(X_i, Y_i)_{i=1}^n$ from the following distribution:

$$X_i \sim \mathcal{N}(0, 1)$$
$$\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$$

and $Y_i = X_i + \varepsilon_i$, where $i = 1, 2, \ldots, n$. We denote the joint distribution of (X, Y) as p.

a) Calculate the population risk $\mathbb{E}_p\mathcal{L}(f(X), Y)$ for a quadratic loss function $\mathcal{L}$, a family of linear functions f with a parameter a, i.e. $f(x) = ax$ and the distribution p. Find the argument that minimises the population risk. How the minimum value changes with σ^2? Implement a function computing the population risk.

b) Let $n = 50$ and $\sigma^2 = 1$. Generate a sample $(X_i, Y_i)_{i=1}^n$. Implement a function computing the empirical risk.

c) Visualise the previous results. Compute the argmin and the min of the population risk and empirical risk. Next, compute the value of the population risk at the point minimising the empirical risk. Mark the three points in the plot. Which point corresponds to the ERM estimator?

---

[1] Recall: $\phi_{\mu,\Sigma}(x) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp(-(x-\mu)'\Sigma^{-1}(x-\mu))$
[2] Recall: $\phi_\lambda(x) = \lambda \exp(-\lambda x)\mathbb{I}(x \geq 0)$

d) Compute the excess risk. Estimate the unconditional excess risk by generating the data $L = 200$ times and averaging excess risks. How the unconditional excess risk changes with a sample size?

e) (*) Do the tasks a)-c) for $X_i \sim \mathcal{N}(\mu_x, \sigma_x^2)$, $Y_i = a_0 X_i + b_0 + \varepsilon_i$ and $f(x) = ax + b$.

f) (*) (estimation error vs approximation error) Do the tasks a)-c) assuming that the true parameter a^* equals 1.1 and we consider a family of linear functions f with a parameter $a \in \mathbb{Z}$.