Deep Learning KU (708.220) WS23

Assignment 1: Maximum Likelihood Estimation

Consider a classification problem with two classes $\mathcal{C}_0$ and $\mathcal{C}_1$. For each class $\mathcal{C}_k$, the samples come from a d-dimensional Gaussian distribution with mean vector $\boldsymbol{\mu}_k$ and a covariance matrix $\boldsymbol{\Sigma}_k = \sigma_k^2 \boldsymbol{I}_d$, where $\boldsymbol{I}_d$ is the $d \times d$ identity matrix and $\sigma_k \in \mathbb{R}^+$.

Probability of data point vector $\boldsymbol{x}$ conditioned on class k equals:

$$p(\boldsymbol{x}|\mathcal{C}_k) = \frac{1}{\sqrt{(2\pi)^d |\boldsymbol{\Sigma}_k|}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu_k})^T \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{x} - \boldsymbol{\mu_k})\right).$$

Hint: $|\boldsymbol{\Sigma}_k|$ is the determinant of $\boldsymbol{\Sigma}_k = \sigma_k^2 \boldsymbol{I}_d$, and equals σ_k^{2d}.

Your training set consists of samples $\boldsymbol{X} = \langle \boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)} \rangle$, where the data points $\boldsymbol{x}^{(m)} \in \mathbb{R}^d$ are independently and identically distributed. You have the corresponding binary targets $\boldsymbol{t} = (t^{(1)}, \ldots, t^{(N)})^T$, with $t^{(m)} \in \{0, 1\}$, which indicates the class of the input sample (i.e., $t^{(m)} = 1$ indicates class $\mathcal{C}_1$).

You will fit a parameterized model for the data-generating distribution:

$$p(\boldsymbol{X}, \boldsymbol{t}|\boldsymbol{\theta}) = p(\boldsymbol{t}|\boldsymbol{\theta}) \cdot p(\boldsymbol{X}|\boldsymbol{t}, \boldsymbol{\theta}).$$

Your model includes a prior probability for the occurrence of each class, where class $\mathcal{C}_0$ occurs with probability $P(\mathcal{C}_0) = p_0$, and class $\mathcal{C}_1$ occurs with probability $P(\mathcal{C}_1) = 1 - p_0$. The parameters of your model are: $\boldsymbol{\theta} = \langle p_0, \boldsymbol{\mu}_0, \boldsymbol{\mu}_1, \sigma_0, \sigma_1 \rangle$.

Task details:

a) (3 pts) : Write the likelihood $p(\boldsymbol{x}^{(m)}, t^{(m)}|\boldsymbol{\theta})$ of a single example $\boldsymbol{x}^{(m)}$, $t^{(m)}$. Accordingly, write the likelihood $p(\boldsymbol{X}, \boldsymbol{t}|\boldsymbol{\theta})$ of the whole training set $\boldsymbol{X}, \boldsymbol{t}$, and then use this to derive the log-likelihood of the training set.

b) (3 pts) : Derive the maximum-likelihood estimate of μ_1 for this model.

c) (2 pts) : Derive the maximum-likelihood estimate of p_0 for this model.

d) (2 pts) : Let's say we are interested in classifying samples by minimizing expected loss, where the loss matrix L will be expressed as:

$$L = \begin{bmatrix} 0 & 20 \\ 1 & 0 \end{bmatrix}.$$

Firstly, using Bayes' rule, express $p(\mathcal{C}_0|\boldsymbol{x})$ and $p(\mathcal{C}_1|\boldsymbol{x})$ in terms of p_0. Then use these to derive an expression for the loss, for each possible classification outcome (i.e., correct $\mathcal{C}_0$, correct $\mathcal{C}_1$, false $\mathcal{C}_0$, false $\mathcal{C}_1$).

Total: 10 points

Provide full derivations including intermediate steps. Present your results clearly, structured and legible.

Assignment details:

- *Assignment issued:* October 11th, 2023, 08:00
- *Deadline:* October 25th, 2023, 08:00
- *Solution submission:* Upload to TeachCenter as one PDF.
- *Rules:* **There are no groups allowed for this task.** Please submit your individual work for your assignment. Copying of solutions or reports from other students is strictly forbidden.