EECE454: Assignment#1

Due: 10.03.2024 23:59PM, via PLMS

Q1. Maximum Likelihood Estimates (20pt)

The Poisson random variable is a discrete variable whose probability mass function is given as

$$P(x|\lambda) = \frac{\lambda^x \cdot \exp(-\lambda)}{x!}, \qquad x \in \{0, 1, 2, \dots, \}, \tag{1}$$

where $\lambda \in (0, \infty)$ is some fixed value that is not yet known to us.
Our goal is to estimate the value of λ from the data. Assume that we have n data points $\{X_1, \dots, X_n\}$, drawn i.i.d. from the Poisson distribution.

Q1.1.

What is the mean of this random variable?

Q1.2.

What is the maximum likelihood estimate (MLE) of λ? In other words, what is the value of λ that maximizes the joint probability of data $X_1, \dots, X_n$ to be drawn from the distribution?

(Hint: This should be a function of the data $X_1, \dots, X_n$.)

Q2. Deriving Naïve Bayes (20pt)

Consider a naïve Bayes binary classifier with the univariate Gaussian likelihood model. That is, we assume that each data point (x, y) has been drawn independently from the joint probability distribution

$$P_{XY}(x, y) = P_{X|Y}(x|y) \cdot P_Y(y), \tag{2}$$

where the likelihood model is

$$P_{X|Y}(x|y) = \mathcal{N}(x|\mu_y, \sigma_y^2), \qquad y \in \{0, 1\} \tag{3}$$

and the prior is the Bernoulli distribution

$$P_Y(0) = 1 - p, \quad P_Y(1) = p. \tag{4}$$

We now assume that we are given the dataset $\{(x_i, y_i)\}_{i=1}^n$.

Q2.1.

Derive the naïve Bayes estimate of p formally.

Q2.2.

Derive the naïve Bayes estimate of $\mu_0, \mu_1, \sigma_0, \sigma_1$ formally.

Q3. Ridge Regression (40pt)

Suppose that we have a regression task at hand. In other words, we have n copies of training data $\{(\mathrm{x}_i, y_i)\}_{i=1}^n$ with $\mathrm{x}_i \in \mathbb{R}_d$ and $y \in \mathbb{R}$, and we want to find a nice function that approximates y. Recall that the **linear regression** typically minimizes the average ℓ_2 loss of linear predictors. To simplify the problem, we ignore the bias and write

$$\min_{\mathrm{w}\in\mathbb{R}^d} \frac{1}{n}\sum_{i=1}(y_i - \mathrm{w}^\top \mathrm{x}_i)^2. \tag{5}$$

What we call the **ridge regression** typically solves the same problem, but with a regularizer. That is, we minimize

$$\min_{\mathrm{w}\in\mathbb{R}^d} \left(\frac{1}{n}\sum_{i=1}(y_i - \mathrm{w}^\top \mathrm{x}_i)^2\right) + \lambda \cdot \|\mathrm{w}\|_2^2, \tag{6}$$

where $\lambda \geqslant 0$ is some hyperparameter.

Q3.1.

Re-express the ridge regression (Eq.6) using $\mathrm{X} = \begin{bmatrix} \mathrm{x}_1^\top \\ \cdots \\ \mathrm{x}_n^\top \end{bmatrix}$ and $\mathrm{y} = \begin{bmatrix} \mathrm{y}_1 \\ \cdots \\ \mathrm{y}_n \end{bmatrix}$, instead of (x_i, y_i).

Q3.2.

Identify the critical point condition of the ridge regression, in terms of X and y.

Q3.3.

Using the critical point condition, find the solution of Eq. 6.

Q3.4.

Under what condition is the critical point of ridge regression unique?

Q4. Optimal predictor (20pt)

Let $X \in \mathbb{R}$ and $Y \in \mathbb{R}$ be continuous random variables, jointly distributed with the density p_{XY}. Suppose that we have a full knowledge about this p_{XY}. We want to find a nice continuous function $f : \mathbb{R} \to \mathbb{R}$ such that the expected value of some loss is minimized, i.e., solve

$$\min_f \mathbb{E}_{(X,Y)\sim p_{XY}}[\ell(f(X), Y)]$$

for some loss function $\ell(\cdot, \cdot)$. Let us call such solution f^*.

Q4.1.

If our loss function is the squared loss, i.e., $\ell(a, b) = (a - b)^2$, show that $f^*(x) = \mathbb{E}[Y|X = x]$.

Q4.2.

If we use the ℓ_1 loss, i.e., $\ell(a, b) = |a - b|$, describe what our f^* should be?