Tsinghua-Berkeley Shenzhen Institute
LEARNING FROM DATA
Fall 2022

Writing Assignment 2

Issued: Saturday 15th October, 2022 **Due:** Saturday 29th October, 2022

POLICIES

- **Acknowledgments:** We expect you to make an honest effort to solve the problems individually. As we sometimes reuse problem set questions from previous years, covered by papers and web pages, we expect the students **NOT** to copy, refer to, or look at the solutions in preparing their answers (relating to an unauthorized material is considered a violation of the honor principle). Similarly, we expect you to not google directly for answers (though you are free to google for knowledge about the topic). If you do happen to use other material, it must be acknowledged here, with a citation on the submitted solution.
- **Required homework submission format:** You can submit homework either as one single PDF document or as handwritten papers. Written homework needs to be provided during the class on the due date, and a PDF document needs to be submitted through Tsinghua's Web Learning (`http://learn.tsinghua.edu.cn/`) before the end of the due date.
- **Collaborators:** In a separate section (before your answers), list the names of all people you collaborated with and for which question(s). If you did the HW entirely on your own, **PLEASE STATE THIS**. Each student must understand, write, and hand in answers of their own.

2.1. **Poisson regression and the exponential family** (3 points)

(a) (1 point) Consider the Poisson distribution parameterized by λ:

$$p(y; \lambda) = \frac{e^{-\lambda}\lambda^y}{y!}.$$

Show that the Poisson distribution is in the exponential family, and clearly state what are $b(y), \eta, T(y)$, and $a(\eta)$.

(b) (1 point) Consider performing regression using a GLM model with a Poisson response variable. What is the canonical response function for the family? (You may use the fact that a Poisson random variable with a parameter λ has mean λ.)

(c) (1 point) For a training set $\{(x^{(i)}, y^{(i)})\}_{i=1}^m$, let the log-likelihood of an example $\log p(y^{(i)}|x^{(i)}; \theta)$. By taking the derivative of the log-likelihood with respect to θ_j, derive the stochastic gradient ascent rule for learning using a GLM model with Poisson responses y and the canonical response function.

2.2. **Gaussian discriminant analysis** (4 points)

Suppose we are given a dataset $\{(\boldsymbol{x}^{(i)}, y^{(i)}) \colon i = 1, 2, \ldots, m\}$ consisting of m independent examples, where $\boldsymbol{x}^{(i)} \in \mathbb{R}^n$ are n-dimension vector, and $y^{(i)} \in \{1, 2, \ldots, k\}$. We will model the joint distribution of $(\boldsymbol{x}, y)$ according to:

$$y^{(i)} \sim \text{Multinomial}(\phi_1, \ldots, \phi_\text{k})$$
$$\boldsymbol{x}^{(i)} | y^{(i)} = j \quad \sim \mathcal{N}(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$$

where the parameter ϕ_j gives $p(y^{(i)} = j)$ for each $j \in \{1, 2, \ldots, k\}$.
In Gaussian Discriminant Analysis (GDA), Linear Discriminant Analysis (LDA) assumes that the classes have a common covariance matrix $\boldsymbol{\Sigma}_j = \boldsymbol{\Sigma}, \forall j$. If the $\boldsymbol{\Sigma}_j$ are not assumed to be equal, we get Quadratic Discriminant Analysis (QDA). The estimates for QDA are similar to those for LDA, except that separate covariance matrices must be estimated for each class. Give the maximum likelihood estimate of $\boldsymbol{\Sigma}_j$ in the case when $k = 2$.

2.3. **Naive Bayes Parameter Learning** (3 points)

Suppose we are given dataset $\{(\boldsymbol{x}^{(i)}, y^{(i)}), i = 1, 2, \ldots, m\}$ consisting of m independent examples, where $\boldsymbol{x}^{(i)} \in \mathbb{R}^n$ are n-dimension vector with entry $\boldsymbol{x}_j \in \{0, 1\}$, and $y^{(i)} \in \{0, 1\}$. We will model the joint distribution of $(\boldsymbol{x}, y)$ according to:

$$y^{(i)} \sim \text{Bernoulli}(\phi_\text{y})$$
$$\boldsymbol{x}_j^{(i)} | y^{(i)} = b \quad \sim \text{Bernoulli}(\phi_{\text{j|y=b}}), \text{b} = 0, 1$$

where the parameters $\phi_y \overset{\text{def}}{=} p(y = 1)$ and $\phi_{j|y=b} \overset{\text{def}}{=} p(\boldsymbol{x}_j = 1 | y^{(i)} = b)$. Under Naive Bayes (NB) assumption, the probability of observing $\boldsymbol{x}_j | y = b, j = 1, \ldots, n$ are independent which means $p(x_1, \cdots, x_n | y) = \Pi_{j=1}^n p(x_j | y)$. Calculate the maximum likelihood estimation of those parameters.