CS480/680: Introduction to Machine Learning

Homework 3

Due: 11:59 pm, June 1, 2020, submit on LEARN.

Include your name and student number!

Submit your writeup in pdf and all source code in a zip file (with proper documentation). Write a script for each programming exercise so that the TAs can easily run and verify your results. Make sure your code runs!
[Text in square brackets are hints that can be ignored.]

Exercise 1: k-NN and Bayes error (6 pts)

Consider the following distribution for feature vector $\mathbf{X} \in \mathbb{R}^d$ and binary label $Y \in \{\pm 1\}$:

$$\mathbf{X} \sim \text{Uniform}([-1,1]^d), \quad \Pr(Y = 1|\mathbf{X} = \mathbf{x}) = \mathsf{sgm}\left(\frac{\mathbf{w}^\top \mathbf{x} + b}{\sigma}\right), \tag{1}$$

where $\mathsf{sgm}(t) := \frac{1}{1+\exp(-t)}$, and $\sigma > 0$ is a parameter that controls the difficulty of our problem. For definiteness, we set $d = 2$, $\mathbf{w} = (1, \underbrace{0, \ldots, 0}_{d-1})$ and $b = 0$ below. For Ex 1.1 - 1.4 we also set $\sigma = 1$.

1. (1 pt) Let us first generate a dataset according to the distribution defined in Equation (1). First, we draw $n = 6000$ samples $\mathbf{X}_i \overset{iid}{\sim} \text{Uniform}([-1,1]^d), i = 1, \ldots, n$. Next, we set $Y_i = 1$ iff $U_i \leq \mathsf{sgm}\left(\frac{\mathbf{w}^\top \mathbf{X}_i + b}{\sigma}\right)$ (otherwise set $Y_i = -1$) where U_i is an independent sample from Uniform([0,1]). Submit your script with suggested format: `(X, y) = data_gen(n, d, sigma, w, b)` where $\mathbf{X} \in \mathbb{R}^{n \times d}$ and $\mathbf{y} \in \{\pm 1\}^n$.

   [You can generate uniform([0,1]) samples using `numpy.random.rand(n, d)`. For definiteness you may use `numpy.random.seed(0)` to fix the random seed.]

2. (2 pt) Compute the Bayes error for the distribution defined in Equation (1). Choose one of the following two options:

   - Refer to Theorem 4.28 in the updated note and compute the error analytically. Submit your derivation.
   - Generate a large enough dataset (as in Ex 1.1) and predict with the Bayes classifier:

     $$\hat{Y}_i = 1 \iff \mathbf{w}^\top \mathbf{X}_i + b \geq 0. \tag{2}$$

     Your prediction error on a large enough dataset will mimic the Bayes error. Report your estimated error and submit your script `error = Bayes(X, y, w, b)`.
     [Hint: How large is large enough? Try different n and see if results change much. If not, then it is probably large enough.]

   In the remaining, we split the generated dataset into $n/2$ training samples and $n/2$ test samples. [Or you may call `data_gen` twice to generate the training set and test set, respectively.]

3. (1 pt) Implement the k-NN classifier and report prediction errors for $k = \{1, 3, 5\}$ on the test set generated as in Ex 1.1. Submit your script `(testYhat, error) = knn(trainX, trainY, testX, k, dist, testY)`. Use the ℓ_2 distance $\mathsf{dist}(\mathbf{x}, \mathbf{x}') = \|\mathbf{x} - \mathbf{x}'\|_2$.

4. (1 pt) Repeat Ex 1.3 with the ℓ_1 distance $\mathsf{dist}(\mathbf{x}, \mathbf{x}') = \|\mathbf{x} - \mathbf{x}'\|_1$, where $\|\mathbf{z}\|_1 = \sum_j |z_j|$.

5. (1 pt) Report the Bayes error and 1-NN error on 4 datasets generated as in Ex 1.1 but with $\sigma = \{0.01, 0.1, 1, 10\}$ respectively. Can you verify the relationship between the two errors? Use the ℓ_2 distance. [No need to split the dataset into train and test when computing the Bayes error.]

Exercise 2: Poisson Regression (4 pts)

Recall that in logistic regression we assumed the binary label $Y_i \in \{0, 1\}$ follows the Bernoulli distribution: $\Pr(Y_i = 1|X_i) = p_i$, where p_i also happens to be the mean. Under the independence assumption we derived the log-likelihood function:

$$\sum_{i=1}^{n}(1-y_i)\log(1-p_i)+y_i\log(p_i). \tag{3}$$

Then, we parameterized the mean parameter p_i through the logit transform:

$$\log\frac{p_i}{1-p_i}=\mathbf{w}^\top\mathbf{x}_i+b, \quad \text{or equivalently} \quad p_i=\frac{1}{1+\exp(-\mathbf{w}^\top\mathbf{x}_i-b)}. \tag{4}$$

Lastly, we found the weight vector $\mathbf{w}$ and b by maximizing the log-likelihood function.

In the following we generalize the above idea to the case where $Y_i \in \mathbb{N}$, i.e., Y_i can take any natural number (for instance, when we are interested in predicting the number of customers or network packages).

1. (1 pt) Naturally, we assume $Y_i \in \mathbb{N}$ follows the Poisson distribution (with mean $\mu_i \geq 0$):

$$\Pr(Y_i=k|X_i)=\frac{\mu_i^k}{k!}\exp(-\mu_i), \quad k=0,1,2,\ldots. \tag{5}$$

   Given a dataset $\mathcal{D}=\{\mathbf{x}_i,y_i\}_{i=1}^n$, what is the log-likelihood function (of μ_i's) given $\mathcal{D}$?

2. (1 pt) Can you give some justification of the parameterization below?

$$\log\mu_i=\mathbf{w}^\top\mathbf{x}_i+b. \tag{6}$$

3. (1 pt) Based on the above, write down the objective function for Poisson regression. Please specify the optimization variables and whether you are maximizing or minimizing. [Constants can be dropped.]

4. (1 pt) Compute the gradient of your objective function above and formulate a gradient algorithm for finding the weight vector $\mathbf{w}$ and b.