Homework Assignment # 2

Due: Friday, February 7, 2025, 11:59 p.m. Mountain time
Total marks: 100

Question 1. [15 MARKS]

Your goal in this question is to find the closed-form solution for the following constrained optimization problem, for fixed non-negative coefficients $c_1, \dots, c_m \geq 0$.

$$\min_{\mathbf{w} \in [0,1]^m, \sum_{k=1}^m w_k = 1} \quad -\sum_{k=1}^{m} c_k \ln w_k$$

(a) [5 MARKS]

What is the Lagrangian L for this optimization? Be clear about what variables are given to the Lagrangian function.

(b) [10 MARKS]

Derive the closer-form solution to the above optimization problem. See Section 6.3, to see the general way to approach this problem. Show your work.

Question 2. [10 MARKS]

Using prototype features provide a simple approach for non-linear data representation. In this question you will reason about some practical choices when using prototype representations.

(a) [5 MARKS]

As a practitioner using kernel regression, a key decision is the choice of kernel similarity. Go to `https://en.wikipedia.org/wiki/Kernel_method#Popular_kernels` and read about one of the linked kernel functions other than RBFs. Provide a 2-3 sentence description of the kernel and what it is used for. **Do not copy and paste from Wikipedia**; explain the kernel in your own words.

(b) [5 MARKS]

You need to decide on how many prototypes m to select. In particular, your goal is to have good generalization performance. Imagine you pick $m = 100$ prototypes out of $n = 1000$ samples. First, do you expect your training error to be lower or higher if you add another 100 prototypes from the dataset? Explain your answer.

Second, do you expect your generalization error to be lower or higher if you add another 100 prototypes from the dataset? Explain your answer.

Question 3. [55 MARKS]

Please complete the Python notebook by doing the following.

(a) [10 MARKS] Complete the parts of the code needed to create a kernel representation.

(b) [10 MARKS] Implement Lasso regression for a linear model.

(c) [10 MARKS] Fill in the necessary implementation details for mini-batch gradient descent using the mean squared error (MSE) loss.

(d) [5 MARKS] Implement the *ConstantLR* optimizer.

(e) [10 MARKS] Implement the prototype selection strategy that uses Lasso regression.

(f) [10 MARKS] Implement (external) cross validation using repeated random subsampling. Note that we do not use internal cross validation in this assignment, and instead simply fix the hyper-parameters to reasonable values that we found.

Question 4. [20 MARKS]

Now you will run two experiments, in the same notebook.

(a) [10 MARKS]

The first experiment uses a synthetic dataset with correlated features and a linear model. The features are generated from a multivariate Gaussian distribution with covariance matrix built using $\Sigma_{ij} = \rho^{|i-j|}$ for some $\rho \in [0, 1.0)$. The weights $\boldsymbol{\beta}$ are randomly generated from a Gaussian distribution and fixed. The targets are generated using $y = \mathbf{x}\boldsymbol{\beta} + \epsilon$, where the noise ϵ is generated from a zero-mean Gaussian distribution with standard deviation $\sigma = 3$.

The goal of this experiment is to compare OLS, Ridge Regression, and Lasso on a dataset with highly correlated features. What conclusions do you draw with the default setting of $\rho = 0.7$? What would you expect to happen as ρ decreases? As ρ increases? Please explain your reasoning. After you make your hypotheses, change ρ and see what you observe.

(b) [10 MARKS]

The second experiment compares different data representations.

1. What conclusions can you draw? Specifically, what do you notice about the difference between the linear and kernel representations.
2. If you were to decide to use a prototype selection strategy in the future, which would you choose? Explain your reasoning. When might we expect the Lasso (ℓ_1) prototype selection strategy to perform better than random selection?

Homework policies:

Your assignment should be submitted on eClass as a single pdf document and a zip file containing: the code and a pdf of written answers. The answers must be written legibly and scanned or must be typed (e.g., Latex).

We will not accept late assignments. Plan for this and aim to submit at least a day early. If you know you will have a problem submitting by the deadline, due to a personal issue that arises, please contact the instructor as early as possible to make a plan. If you have an emergency that prevents submission near the deadline, please contact the instructor right away. Retroactive reasons for delays are much harder to deal with in a fair way.

All assignments are individual. All the sources used for the problem solution must be acknowledged, e.g. web sites, books, research papers, personal communication with people, etc. Academic honesty is taken seriously; for detailed information see the University of Alberta Code of Student Behaviour.

Good luck!