CS559 Machine Learning: Fundamentals and Applications
Fall 2022 HW#2
Due: 10/7/2022 11:59 PM

- The assignment must be individual work and must not be copied or shared. Any tendency of cheating/copying evidence will lead to a 0 mark for the assignment.
- There are two types of problems – conceptual paper-based and application notebook-based. The conceptual paper-based problems need to be worked on paper and submitted in a single pdf file. Please name the file as *Name_HW#.pdf*.
- The application problems must be worked on in the notebook using python. If the problem does not mention the libraries/packages to use, students must only use pandas, numpy, and spacy. All problems must be in a single notebook file as *Name_HW#.ipynb*.
- All files must be compressed into a single zip file as *Name_HW#.zip*.

Question 1 [10 points]: Binary Classification

Consider a situation that the common distribution used in the binary classification, $\mu \in \{0,1\}$, forms the Bernoulli distribution

$$p(x|\mu) = \mu^x(1-\mu)^{1-x}$$

not symmetric between two values of *x*. In this case, it is much convenient to use an equivalent formulation with $\mu \in \{-1,1\}$ instead. The probability distribution then can be expressed as follow,

$$p(x|\mu) = \left(\frac{1-\mu}{2}\right)^{\frac{1-x}{2}} \left(\frac{1+\mu}{2}\right)^{\frac{1+x}{2}}.$$

Show that the distribution is normalized and evaluate its mean and variance.

Question 2 [10 points]: Bayesian Theorem

Consider a *D*-dimensional Gaussian random variable x with distribution $\mathcal{N}(x|\mu,\Sigma)$ in which the covariance Σ is known and for which we wish to infer the mean μ from a set of observations $X = \{x_1, \ldots, x_N\}$. Given a prior distribution $p(\mu) = \mathcal{N}(\mu|\mu_0, \Sigma_0)$, find the corresponding posterior distribution $p(\mu|X)$.

Question 3 [60 points]: Linear Classifier Implementations

a) [35 points: 15 points for each model, 5 points for evaluation] Using the provided dataset in the notebook, implement the perceptron and linear discriminant analysis (LDA) algorithms (use NumPy, Pandas) to classify the dataset. Then, report the accuracy of trained models.
b) [5 points] Use the provided test dataset, general and evaluate the models.
c) [15 pts] Implement models (perceptron, LDA, and logistic regression) using Scikit-learn.
d) [5 points] Compare the weights obtained and the accuracies of trained models. Did you expect weights to be the same? Explain why or why not?

Question 4 [20 points]: Linear Regression and regularization

Suppose we decided to reject the linear model for the housing dataset discussed in linear regression lecture (the second model with all features after the preprocessing). Then, we are going to implement a new model that is less complex (lower number of features) and outperform (lower RMSE).

a) [10 points] Retrain the dataset. In the new trained dataset, same feature engineering for the continuous features is not allowed while the discrete features may be used as they are. For each process, explain explicitly the reason of work.
b) [10 points] Suppose the worst scenario is when the RMSE of new model is higher but within 5% with the same number of features that the original model had. Train five different models (with or without regularizations) and report the results. Select the best model among five models and explain why.