CPSC/MATH 499/699 Assignment 1

Due online by 6pm on October 9

General Instructions: Submit online a PDF copy of your written solution with your name and student number. Be sure scanned pages are in the correct orientation, with written answers for each questions clearly labelled. For questions involving programming, submit a Jupyter notebook with all relevant program code, outputs, comments and explanations of your results. You may work together but you must write your own solution.
Online Submission Files: A PDF file named A1-StudentNumber-LastName-FirstName.pdf and a Jupyter Notebook named A1-StudentNumber-LastName-FirstName.ipynb. For example, my submission files would be named `A1-123456789-Wan-Andy.pdf` and `A1-123456789-Wan-Andy.ipynb`.

For CPSC/MATH 499, answer Q1, Q2 and either Q3 or Q4.
For CPSC/MATH 699, answer all 4 questions.

Q1. Introduction to Machine Learning
(a) Suppose you wish to develop a model to predict your letter grade in a course. What kind of machine learning problem might this involve and what might some relevant features and data be?
(b) Would you perform batch or online learning for a weather prediction model? Explain.
(c) Explain how you can use clustering to generate more labelled data for semi-supervised learning.
(d) Formulate the problem of detecting spam emails as supervised learning. What might the data be and what machine learning tasks might you need to solve to make the detection work? Can you also solve this problem with semi-supervised or unsupervised learning? Explain.

Q2. Adding more features to the used car price prediction model
(a) From the class demo code and using the fordFocus.csv data with all colors, train a linear regression model with features on year and mileage to predict the price.
Hint: To use both features, add `features = ['year', 'mileage']` *before you train your model.*
Split your data set randomly into 80% for training and 20% for testing. Plot the data and your resulting model on price versus year and mileage. Compute the R^2 value and MSE using test data and compare them against a linear regression model using mileage only. Is the mileage and year model performing better? Explain why or why not.
(b) Next train a polynomial regression model with variable degree on year and mileage. Plot the resulting MSE for each degree using the test data. For what degree(s) is the model underfitting and overfitting?

Q3. Relationship between Bias and Variance of the expected mean squared error (noisy case)
(a) Recall the linearity of expectation of two random variables,

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y], \text{ for } a, b \in \mathbb{R}.$$

Use the linearity of expectation to show the following three properties about expectation, variance and covariance.

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \tag{1}$$

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y] + \mathrm{cov}(X, Y) \tag{2}$$

$$\mathrm{Var}(X \pm Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) \pm 2\mathrm{cov}(X, Y) \tag{3}$$

(b) Assume data (x, y) is randomly drawn from a distribution D, with m training samples S from D^m and noise ϵ from another distribution E. Further suppose that the output is related by $y = f(x) + \epsilon$ for some unknown function f, and the predicted output using S training data is $\hat{y} = \hat{f}_S(x)$. The expected mean square error with noise is defined as $\mathbb{E}[(y - \hat{y})^2]$, where the expectation is taken over all $(x, y) \sim D, S \sim D^m, \epsilon \sim E$.
Suppose the noise has zero mean and is independent from $y, \hat{y}$ (i.e. $\mathbb{E}[\epsilon] = 0, \mathrm{cov}(y, \epsilon) = 0, \mathrm{cov}(\hat{y}, \epsilon) = 0$), then show that

$$\mathbb{E}[(y - \hat{y})^2] = (\mathbb{E}[f(x)] - \mathbb{E}[\hat{f}_S(x)])^2 + \mathrm{Var}(f(x) - \hat{f}_S(x)) + \mathrm{Var}(\epsilon).$$

Hint: Follow similar steps from class for the noiseless case and apply clearly each properties from part **(a)**.

Q4. Singular Value Decomposition

(a) Following the proof of existence of SVD, compute by hand the SVD of $A = \begin{pmatrix} 2 & 0 \\ 1 & 1 \\ 0 & 2 \end{pmatrix}$. Verify that your U, Σ, V indeed satisfy $A = U\Sigma V^T$. *Hint: First diagonalize A^TA to get the eigenvectors $\boldsymbol{v}_i$ and eigenvalues λ_i. Then set $\sigma_i = \sqrt{\lambda_i}$ and find the remaining $\boldsymbol{u}_i$ as in the proof.*
In the following, let A be a $m \times n$ real matrix with rank $r \leq \min\{m, n\}$ and $A = U\Sigma V^T$ be its singular value decomposition with singular values denoted as σ_i.
(b) Supposing $m = n = r$ (A is an invertible $n \times n$ matrix), show that the solution to $A\boldsymbol{x} = \boldsymbol{b}$ can be given by $\boldsymbol{x} = V\Sigma^{-1}U^T\boldsymbol{b}$. Explain why one may not want to use this in practice.
(c) Supposing $r = n$ and $m \geq n$, recall the least square solution to $A\boldsymbol{\theta} = \boldsymbol{y}$ is given by $\boldsymbol{\theta} = (A^TA)^{-1}A^T\boldsymbol{y}$. Using SVD, show that it can expressed as $\boldsymbol{\theta} = V\Sigma^{-1}U^T\boldsymbol{y}$. Compare with the expression derived in class.
(d) Recall the Frobenius norm of A is $\|A\|_F^2 = \mathrm{tr}(A^TA)$. Show that $\|A\|_F^2 = \sigma_1^2 + \cdots + \sigma_r^2$.