Homework 1
CS534 Machine Learning, Spring 2019

This homework explores the concepts of math with random variables, covariance, and linear regression. Points are noted in each problem. There are no time limits.

Problem 1 - Expectations (10 points)

Given a model of a random process

$$Y = f(X) + \epsilon \tag{1}$$

where Y is what is measured, X are variables, and $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$ is gaussian noise.

1.a. Suppose you have a prediction model $\hat{f}$, show that the expected error $Err(x) = E[(Y - \hat{f}(x))^2 | X = x]$ can be written as

$$Err(x) = \sigma_\epsilon^2 + (E[\hat{f}(x)] - f(x))^2 + E[(\hat{f}(x) - E[\hat{f}(x)])^2] \tag{2}$$

1.b. Describe what each of these terms represents in plain English.

Problem 2 - Covariance (10 points)

2.a. Suppose you want to generate samples from a normally-distributed random variable $X \sim \mathcal{N}(\mu, \Sigma), X \in \mathbb{R}^p$. Show with math that you can transform samples from the standard normal distribution $\mathcal{N}(0, I)$ (where I is the identity) to match this distribution using the diagonalization $\Sigma = V \Lambda V^T$.

2.b. Use this procedure to generate $N = 1000$ samples from the distribution

$$X = [x_1, x_2]^T \sim \mathcal{N}(\mu, \Sigma), \quad \mu = [1, 1]^T, \quad \Sigma = \begin{bmatrix} 1 & -0.5 \\ -0.5 & 0.5 \end{bmatrix}. \tag{3}$$

Display the samples x using a scatter plot. Superimpose the eigenvectors and the level curves of the PDF on the scatter plot.

2.c. In a new figure superimpose the level curves of the Euclidean distance on the scatter plot

$$D(x) = \|x - \mu\|_2 \tag{4}$$

2.d. In a new figure superimpose the level curves of the *Mahalanobis distance* on the scatter plot

$$M(x) = \sqrt{(x - \mu)^T \Sigma^{-1} (x - \mu)} \tag{5}$$

Problem 3 - Regularizing linear regression (15 points)

In this problem we have a dataset that contains $p = 100$ features, but the underlying model that relates X, Y involves only 5 of these

$$Y = \beta_{j_1} X_{j_1} + \beta_{j_2} X_{j_2} + \ldots + \beta_{j_5} X_{j_5} + \epsilon \tag{6}$$

for some $j_1, j_2, \ldots, j_3 \in \{1, \ldots, 100\}$.

LASSO regression incorporates a model penalty in the loss function that effectively encourages a *sparse* solution, forcing many model weights β_j towards zero

$$Loss(\beta) = \sum_{i=1}^{N} (y_i - \beta x_i)^2 + \lambda |\beta|_1. \tag{7}$$

Since the L_1 norm is not differentiable, LASSO models can be derived using the sub-gradient method for gradient-descent. This sub-gradient method represents the gradient of the penalty term using a *soft thresholding* operation

$$S(\lambda, \beta_j) = \begin{cases} \beta_j - \lambda & \text{if } \beta_j > \lambda \\ 0 & \text{if } -\lambda \leq \beta_j \leq \lambda \\ \beta_j + \lambda & \text{if } \beta_j < -\lambda. \end{cases} \tag{8}$$

This sub-gradient term is combined with the gradient of the cost term $\sum_{i=1}^{N} (y_i - \beta x_i)^2$ for gradient descent.

3.a. Find a LASSO model $\hat{\beta}$ to the training data using gradient descent. Use the penalty weight $\lambda = 1$ and learning rate $\gamma = 5e-3$ to train the model for 10000 gradient updates. Plot the final model coefficients. Can you guess the indices of the nonzero model weights $\{j_1, \ldots, j_5\}$?

3.b. Fit a Least Squares model without regularization to the training set. Plot the final model coefficients.

3.c. Compare the mean-square error of the LASSO and ordinary least squares models on the testing set.

Problem 4 - Robust fitting with outliers (15 points)

The *RANdom SAmple Consensus* (RANSAC) algorithm has been used for over 38 years to fit models in the presence of large number of outliers. In this problem you will be using data generated from the process

$$Y = f(X) = -3.2591X^3 + 4.8439X^2 + 1.7046X + 1.0685 + \epsilon \qquad (9)$$

where $\epsilon \sim \mathcal{N}(0, 1e-2)$. These samples contain outliers generated from a uniform distribution.

4.a. Use polynomial least squares to estimate $\hat{f}$. Display the samples in a scatter plot and superimpose the true and estimated model on this plot.

4.b. Implement RANSAC to estimate $\hat{f}$. Choose 5 points at random to fit each model, and use the threshold $|y - \hat{f}(x)| \leq 0.3$ to define the consensus set. Stop when the number of inliers exceeds %40 of the total samples, and recalculate the final model using this consensus set. Display the samples in a scatter plot and indicate the final consensus set. Superimpose f, $\hat{f}$ on this plot.