# Assignment #1
Due: 23:55pm September 28, 2018
Upload at: https://dropbox.cs.princeton.edu/COS324_F2018/HW1

## Refreshing Probability and Calculus

### Problem 1 (5pts)
$X$ and $Y$ are two independent random variables with distributions $p_X(x)$ and $p_Y(y)$, respectively. Show that the independence of $X$ and $Y$ implies that their covariance is zero.

### Problem 2 (5pts)
Let $X \in \{0, 1\}$ be a binary random variable with a Bernoulli distribution. Suppose $p(X = 1) = \theta$, where $\theta$ is $0 < \theta < 1$ and $\theta \in \mathbb{R}$. The likelihood of observing a sample $x \sim X$ is

$$p(x \mid \theta) = \theta^x(1 - \theta)^{1-x}.$$  

Prove that $\mathbb{E}[X] = \theta$ and $\text{var}[X] = \theta(1 - \theta)$.

### Problem 3 (9pts)
Let $A, B, C \in \mathbb{R}^{n \times n}$ be invertible symmetric matrices and $x \in \mathbb{R}^n$. We further suppose that all these matrices commute with each other.

A. Show that if $M, N \in \mathbb{R}^{n \times n}$ are two symmetric matrices that commute then $MN$ is also a symmetric matrix.

Compute the gradients of the following quantities with respect to $x$.

B. $x^T Ax$

C. $\text{tr}(Axx^TB)$

D. $\exp\{-(Ax - b)^T C^{-1}x\}$

### Problem 4 (6pts)
Let $X$ be a Gaussian random variable with mean $\mu$ and variance $\sigma^2$. The probability density function for $X$ is

$$p_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x - \mu)^2\right\}.$$  

We now transform $X$ into a variable $Y = e^X$.

A. What is the PDF for $Y$, $p_Y(y)$?

B. What is the mean of $Y$?
Least Squares Regression

**Problem 5** (15pts)
Prove that the least squares objective function
\[ L(w) = (Xw - y)^T (Xw - y) \]
is convex with respect to \( w \).

**Problem 6** (25pts)
When fitting a machine learning model, it may happen sometimes that we wish to weight some data more than others when we fit parameters. Consider a weighted data set \( \{x_n, y_n, r_n\}_{n=1}^N \), where \( x_n \in \mathbb{R}^D \) (with the constant 1 column already included), \( y_n \in \mathbb{R} \), and \( r_n > 0 \). Here the \( r_n \) are weights for each of the data and we’d like to perform least-squares regression accounting for these weights with the following loss:
\[ L(w) = \frac{1}{\sum_{n=1}^N r_n} \sum_{n=1}^N r_n (x_n^T w - y_n)^2. \]
Dividing by the sum here is generalizing the division by \( N \) in the unweighted case. Derive a closed-form solution for \( w \) that minimizes this loss. Hint: it will be helpful to construct a diagonal matrix \( R \) such that \( R_{n,n} = r_n \).

Maximum Likelihood Linear Regression

**Problem 7** (35pts)
Here are some simple data to regress:
\[ x = [-1.87, -1.76, -1.67, -1.22, -0.07, 0.11, 0.67, 1.60, 2.22, 2.51] \]
\[ y = [0.06, 1.67, 0.54, -1.45, -0.18, -0.67, 0.92, 2.95, 5.13, 5.18] \]
For each of the four feature representations below, use maximum likelihood estimation to fit the regression weights and the variance. For each, 1) report the MLE regression weights and variance, and 2) make a plot that shows the data, the resulting predictive mean, 2\( \sigma \) predictive bands around the mean. Which of the representations seems to fit the data best? Explain your reasoning. Turn in the Python file that produces these plots.
A. \( \Phi(x) = [x, 1] \)
B. \( \Phi(x) = [x^2, x, 1] \)
C. \( \Phi(x) = [\sin(3x), \sin(2x), \sin(x), 1] \)
D. \( \Phi(x) = [e^{-(x-2)^2}, e^{-(x-1)^2}, e^{-x^2}, e^{-(x+1)^2}, e^{-(x+2)^2}, 1] \)