Assignment #1

Due: 23:55pm February 18, 2019

Upload at: `https://dropbox.cs.princeton.edu/COS324_S2019/HW1`

Problem 1 (5pt)

In COS 324 you'll typeset your assignments in LaTeX, which is the way that almost everything is composed in mathematical fields. It's a good thing to learn and it can produce beautiful documents. There are a variety of ways to write and compile LaTeX, from web-based tools like Overleaf, to various GUI tools for Mac, Linux and Windows, to plain text editors like Emacs. In this problem, figure out how to compile this document to a PDF in such a way that it shows your name an email in the top left, following the instructions in the comments above. Note that compiling this document does not automatically upload it or anything; when you're done you will still need to upload the PDF (and your code) to the CS dropbox URL at the top of the page.

Refreshing Probability and Calculus

Problem 2 (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 3 (5pts)

Let $X \in \{0, 1\}$ be a binary random variable with a Bernoulli distribution. Suppose $p(X = 1) = \theta$, where θ is $0 < \theta < 1$ and $\theta \in \mathbb{R}$. That is, the probability of $X = x$ given θ is

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

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

Problem 4 (9pts)

Let $A, B, C \in \mathbb{R}^{n \times n}$ be invertible symmetric matrices and $\boldsymbol{x} \in \mathbb{R}^n$. 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 $\boldsymbol{x}$.
B. $\boldsymbol{x}^T A \boldsymbol{x}$
C. $\mathrm{tr}(A \boldsymbol{x} \boldsymbol{x}^T B)$
D. $\exp\{-(A\boldsymbol{x} - \boldsymbol{b})^T C^{-1} \boldsymbol{x}\}$

Problem 5 (6pts)

Let X be a Gaussian random variable with mean μ and variance σ^2. The probability density function for X is

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

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 6 (15pts)

Prove that the least squares objective function

$$L(\mathbf{w}) = (\mathbf{X}\mathbf{w} - \mathbf{y})^T (\mathbf{X}\mathbf{w} - \mathbf{y})$$

is convex with respect to $\mathbf{w}$.

Problem 7 (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 $\{\mathbf{x}_n, y_n, r_n\}_{n=1}^N$, where $\mathbf{x}_n \in \mathbb{R}^D$ (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:

$$L(\mathbf{w}) = \frac{1}{\sum_{n=1}^N r_n} \sum_{n=1}^N r_n (\mathbf{x}_n^T \mathbf{w} - y_n)^2.$$

Dividing by the sum here is generalizing the division by N in the unweighted case. Derive a closed-form solution for $\mathbf{w}$ that minimizes this loss. Hint: it will be helpful to construct a diagonal matrix $\mathbf{R}$ such that $R_{n,n} = r_n$.

Maximum Likelihood Linear Regression

Problem 8 (30pts)

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, 2σ predictive bands around the mean. (That is plot a line 2σ above the mean and another 2σ below 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. $\Phi(x) = [e^{-(x-2)^2}, e^{-(x-1)^2}, e^{-x^2}, e^{-(x+1)^2}, e^{-(x+2)^2}, 1]$

Changelog

- 11 February 2019 – Extension of due date to 18 February.

- 6 February 2019 – Initial version.