

# CS761 Spring 2015 Homework 2

Assigned Mar. 13, due Mar. 27 before class

## Instructions:

- Homeworks are to be done individually.
- Typeset your homework in latex using this file as template (e.g. use `pdflatex`). Show your derivations.
- Hand in the compiled pdf (not the latex file) online. Instructions will be provided. We do not accept hand-written homeworks.
- Homework will no longer be accepted once the lecture starts.
- Fill in your name and email below.

Name:

Email:

1. Let  $X_0, X_1, \dots, X_{M-1}$  denote a random sample of  $N$ -dimensional random vectors  $X_n$ , each of which has mean value  $m$  and covariance matrix  $R$ . Show that the sample mean

$$\hat{m}_t = \frac{1}{t+1} \sum_{n=0}^t X_n$$

and the sample covariance

$$S_t(\hat{m}_t) = \frac{1}{t+1} \sum_{n=0}^t (X_n - \hat{m}_t)(X_n - \hat{m}_t)^\top$$

may be written recursively as

$$\hat{m}_t = \frac{t}{t+1} \hat{m}_{t-1} + \frac{1}{t+1} X_t, \quad \hat{m}_0 = X_0,$$

and

$$S_t(\hat{m}_t) = Q_t - \hat{m}_t \hat{m}_t^\top,$$

where

$$Q_t = \frac{t}{t+1} Q_{t-1} + \frac{1}{t+1} X_t X_t^\top.$$

2. Suppose we roll a fair 6-sided die 100 times. Let  $X$  be the sum of the outcomes. Bound  $P(|X - 350| \geq 100)$  using Chebyshev and Hoeffding, respectively.
3. Let  $\mathcal{X}$  be the vector space of *finitely* nonzero sequences  $X = (x_1, x_2, \dots, x_n, 0, 0, \dots)$ . Define the norm on  $\mathcal{X}$  as  $\|X\| = \max |x_i|$ . Let  $X_n$  be a point in  $\mathcal{X}$  (a sequence) defined by

$$X_n = \left( 1, \frac{1}{2}, \frac{1}{3}, \dots, \frac{1}{n}, 0, 0, \dots \right).$$

- Show that the sequence  $X_n$  is a Cauchy sequence.
- Show that  $\mathcal{X}$  is not complete.

4. Determine the range and nullspace of the following linear operators (matrices):

$$A = \begin{bmatrix} 1 & 0 \\ 5 & 4 \\ 2 & 4 \end{bmatrix} \quad B = \begin{bmatrix} 1 & 0 & 1 \\ 5 & 4 & 9 \\ 2 & 4 & 6 \end{bmatrix}$$

5. Let

$$A = \begin{bmatrix} 1 & 4 & 5 & 6 \\ 6 & 7 & 2 & 1 \end{bmatrix} \quad b = \begin{bmatrix} 48 \\ 30 \end{bmatrix}.$$

One solution to  $Ax = b$  is  $x = [1, 2, 3, 4]^\top$ . Compute the least-squares solution using the SVD (explain how), and compare. Why was the solution chosen?

6. Consider the following process. A probability vector  $p = (p_1, \dots, p_d)$  is drawn from a Dirichlet distribution with parameter vector  $\alpha$ . Then, a vector of category counts  $x = (x_1, \dots, x_d)$  is drawn from a multinomial distribution with probability vector  $p$  and number of trials  $N$ . Give an analytic form of  $P(x | \alpha)$ .

7. Let  $X_1, X_2, \dots, X_m$  be a random sample, where  $X_i \sim U(0, \theta)$  the uniform distribution.

- Show that  $\hat{\theta}_{ML} = \max X_i$ .
- Show that the density of  $\hat{\theta}_{ML}$  is  $f_{\theta}(x) = \frac{m}{\theta^m} x^{m-1}$ .
- Find the expected value of  $\hat{\theta}_{ML}$ .
- Find the variance of  $\hat{\theta}_{ML}$ .

8. Let  $X_1, \dots, X_n$  be a sample from  $N(\mu, \sigma^2)$ .

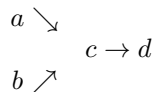
- Show that the MLE of  $\sigma^2$  is

$$\hat{\sigma}^2 = n^{-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

- Show that  $\hat{\sigma}^2$  has a smaller mean squared error than

$$(n-1)^{-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

9. Consider the directed graphical model in which none of the variables is observed.



Show that  $a \perp b | \emptyset$  by using a probability argument. Suppose we now observe the variable  $d$ . Show that in general  $a \not\perp b | d$  (you can use a counterexample).

10. Consider two discrete random variables  $x, y \in \{A, B, C\}$ . Construct a joint distribution  $p(x, y)$  with the following properties:

- $\hat{x}$  is the maximizer of the marginal  $p(x)$
- $\hat{y}$  is the maximizer of the marginal  $p(y)$
- $p(\hat{x}, \hat{y}) = 0$ .

11. Logistic regression for  $y \in \{-1, 1\}$  is defined by

$$p(y | x; w, b) = \frac{1}{1 + e^{-y(x^\top w + b)}}.$$

Show that logistic regression is in the exponential family, that is, the probability distribution can be written in the form

$$p(y | x; \tilde{w}) = \frac{1}{Z(x, \tilde{w})} e^{\phi(y, x)^\top \tilde{w}}.$$

Note the mapping  $\phi$  depends only on  $y, x$ , but not on  $w$  or  $b$ .