

**Homework #0**

RELEASE DATE: 09/13/2015

DUE DATE: NONE

**1 Probability and Statistics**

(1) (combinatorics)

Let  $C(N, K) = 1$  for  $K = 0$  or  $K = N$ , and  $C(N, K) = C(N - 1, K) + C(N - 1, K - 1)$  for  $N \geq 1$ .  
 Prove that  $C(N, K) = \frac{N!}{K!(N-K)!}$  for  $N \geq 1$  and  $0 \leq K \leq N$ .

(2) (counting)

What is the probability of getting exactly 4 heads when flipping 10 fair coins?

What is the probability of getting a full house (XXXYY) when randomly drawing 5 cards out of a deck of 52 cards?

(3) (conditional probability)

If your friend flipped a fair coin three times, and tell you that one of the tosses resulted in head, what is the probability that all three tosses resulted in heads?

(4) (Bayes theorem)

A program selects a random integer  $X$  like this: a random bit is first generated uniformly. If the bit is 0,  $X$  is drawn uniformly from  $\{0, 1, \dots, 7\}$ ; otherwise,  $X$  is drawn uniformly from  $\{0, -1, -2, -3\}$ .  
 If we get an  $X$  from the program with  $|X| = 1$ , what is the probability that  $X$  is negative?

(5) (union/intersection)

If  $P(A) = 0.3$  and  $P(B) = 0.4$ ,  
 what is the maximum possible value of  $P(A \cap B)$ ?  
 what is the minimum possible value of  $P(A \cap B)$ ?  
 what is the maximum possible value of  $P(A \cup B)$ ?  
 what is the minimum possible value of  $P(A \cup B)$ ?

(6) (mean/variance)

Let mean  $\bar{X} = \frac{1}{N} \sum_{n=1}^N X_n$  and variance  $\sigma_X^2 = \frac{1}{N-1} \sum_{n=1}^N (X_n - \bar{X})^2$ . Prove that

$$\sigma_X^2 = \frac{N}{N-1} \left( \frac{1}{N} \sum_{n=1}^N X_n^2 - \bar{X}^2 \right).$$

(7) (Gaussian distribution)

If  $X_1$  and  $X_2$  are independent random variables, where  $p(X_1)$  is Gaussian with mean 2 and variance 1,  $p(X_2)$  is Gaussian with mean  $-3$  and variance 4. Let  $Z = X_1 + X_2$ . Prove  $p(Z)$  is Gaussian, and determine its mean and variance.

**2 Linear Algebra**

(1) (rank)

What is the rank of  $\begin{pmatrix} 1 & 2 & 1 \\ 1 & 0 & 3 \\ 1 & 1 & 2 \end{pmatrix}$ ?

(2) (inverse)

What is the inverse of  $\begin{pmatrix} 0 & 2 & 4 \\ 2 & 4 & 2 \\ 3 & 3 & 1 \end{pmatrix}$ ?

(3) (eigenvalues/eigenvectors)

What are the eigenvalues and eigenvectors of  $\begin{pmatrix} 3 & 1 & 1 \\ 2 & 4 & 2 \\ -1 & -1 & 1 \end{pmatrix}$ ?

(4) (singular value decomposition)

(a) For a real matrix  $M$ , let  $M = U\Sigma V^T$  be its singular value decomposition. Define  $M^\dagger = V\Sigma^\dagger U^T$ , where  $\Sigma^\dagger[i][j] = \frac{1}{\Sigma[i][j]}$  when  $\Sigma[i][j]$  is nonzero, and 0 otherwise. Prove that  $MM^\dagger M = M$ .

(b) If  $M$  is invertible, prove that  $M^\dagger = M^{-1}$ .

(5) (PD/PSD)

A symmetric real matrix  $A$  is positive definite (PD) iff  $\mathbf{x}^T A \mathbf{x} > 0$  for all  $\mathbf{x} \neq \mathbf{0}$ , and positive semi-definite (PSD) if “ $>$ ” is changed to “ $\geq$ ”. Prove:

(a) For any real matrix  $Z$ ,  $ZZ^T$  is PSD.

(b) A symmetric  $A$  is PD iff all eigenvalues of  $A$  are strictly positive.

(6) (inner product)

Consider  $\mathbf{x} \in R^d$  and some  $\mathbf{u} \in R^d$  with  $\|\mathbf{u}\| = 1$ .

What is the maximum value of  $\mathbf{u}^T \mathbf{x}$ ? What  $\mathbf{u}$  results in the maximum value?

What is the minimum value of  $\mathbf{u}^T \mathbf{x}$ ? What  $\mathbf{u}$  results in the minimum value?

What is the minimum value of  $|\mathbf{u}^T \mathbf{x}|$ ? What  $\mathbf{u}$  results in the minimum value?

(7) (distance)

Consider two parallel hyperplanes in  $R^d$ :

$$H_1 : \mathbf{w}^T \mathbf{x} = +3,$$

$$H_2 : \mathbf{w}^T \mathbf{x} = -2,$$

where  $\mathbf{w}$  is the norm vector. What is the distance between  $H_1$  and  $H_2$ ?

### 3 Calculus

(1) (differential and partial differential)

Let  $f(x) = \ln(1 + e^{-2x})$ . What is  $\frac{df(x)}{dx}$ ? Let  $g(x, y) = e^x + e^{2y} + e^{3xy^2}$ . What is  $\frac{\partial g(x, y)}{\partial y}$ ?

(2) (chain rule)

Let  $f(x, y) = xy$ ,  $x(u, v) = \cos(u + v)$ ,  $y(u, v) = \sin(u - v)$ . What is  $\frac{\partial f}{\partial v}$ ?

(3) (integral)

What is  $\int_5^{10} \frac{2}{x-3} dx$ ?

(4) (gradient and Hessian)

Let  $E(u, v) = (ue^v - 2ve^{-u})^2$ . Calculate the gradient  $\nabla E$  and the Hessian  $\nabla^2 E$  at  $u = 1$  and  $v = 1$ .

(5) (Taylor's expansion)

Let  $E(u, v) = (ue^v - 2ve^{-u})^2$ . Write down the second-order Taylor's expansion of  $E$  around  $u = 1$  and  $v = 1$ .

(6) (optimization)

For some given  $A > 0, B > 0$ , solve

$$\min_{\alpha} Ae^{\alpha} + Be^{-2\alpha}.$$

(7) (vector calculus)

Let  $\mathbf{w}$  be a vector in  $R^d$  and  $E(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T\mathbf{A}\mathbf{w} + \mathbf{b}^T\mathbf{w}$  for some symmetric matrix  $\mathbf{A}$  and vector  $\mathbf{b}$ . Prove that the gradient  $\nabla E(\mathbf{w}) = \mathbf{A}\mathbf{w} + \mathbf{b}$  and the Hessian  $\nabla^2 E(\mathbf{w}) = \mathbf{A}$ .

(8) (quadratic programming)

Following the previous question, if  $\mathbf{A}$  is not only symmetric but also positive definite (PD), prove that the solution of  $\operatorname{argmin}_{\mathbf{w}} E(\mathbf{w})$  is  $-\mathbf{A}^{-1}\mathbf{b}$ .

(9) (optimization with linear constraint)

Consider

$$\min_{w_1, w_2, w_3} \frac{1}{2}(w_1^2 + 2w_2^2 + 3w_3^2) \text{ subject to } w_1 + w_2 + w_3 = 11.$$

Refresh your memory on “Lagrange multipliers” and show that the optimal solution must happen on  $w_1 = \lambda$ ,  $2w_2 = \lambda$ ,  $3w_3 = \lambda$ . Use the property to solve the problem.

(10) (optimization with linear constraints)

Let  $\mathbf{w}$  be a vector in  $R^d$  and  $E(\mathbf{w})$  be a convex differentiable function of  $\mathbf{w}$ . Prove that the optimal solution to

$$\min_{\mathbf{w}} E(\mathbf{w}) \text{ subject to } \mathbf{A}\mathbf{w} + \mathbf{b} = \mathbf{0}.$$

must happen at  $\nabla E(\mathbf{w}) + \boldsymbol{\lambda}^T \mathbf{A} = \mathbf{0}$  for some vector  $\boldsymbol{\lambda}$ . (*Hint: If not, let  $\mathbf{u}$  be the residual when projecting  $\nabla E(\mathbf{w})$  to the span of the rows of  $\mathbf{A}$ . Show that for some very small  $\eta$ ,  $\mathbf{w} - \eta \cdot \mathbf{u}$  is a feasible solution that improves  $E$ .)*