CS57300 Homework	— Homework 1 , Problem 1 —	(January 23, 2016)
Xilun Wu (Wu636)		

Q1:

(a) Show that the variance of var(X+Y) = var(X) + var(Y) + 2 cov(X,Y), where cov is the covariance between X and Y.
(b) State one condition over X and Y that makes $\frac{cov(X,Y)}{\sqrt{var(X)}\sqrt{var(Y)}}$ achieve its maximum value.
(c) State one condition over X and Y that makes $\frac{cov(X,Y)}{\sqrt{var(X)}\sqrt{var(Y)}}$ achieve its minimum value.

A:
(a)

$$\begin{aligned} var(X+Y) &= E((X+Y)^2) - E^2(X+Y) && (1)\\ &= (E(X^2) + E(Y^2) + 2E(XY)) - (E(X) + E(Y))^2 && (2)\\ &= (E(X^2) - E^2(X)) + (E(Y^2) - E^2(Y)) + 2(E(XY) - E(X)E(Y)) && (3)\\ &= var(X) + var(Y) + 2cov(X,Y) && (4) \end{aligned}$$

(b)

$$cov(X,Y) = E(XY) - E(X)E(Y) \quad (5)$$

According to (5), if X = kY (k > 0), we have the maximum cov(X,Y):

$$cov(X,Y) = E(XY) - E(X)E(Y) = kE(Y^2) - kE^2(Y) = k * var(Y) \quad (6)$$

$$\begin{aligned} \frac{cov(X,Y)}{\sqrt{var(X)}\sqrt{var(Y)}} &= \frac{k * var(Y)}{k * var(Y)} && (7)\\ &= 1 && (8) \end{aligned}$$

(c)
According to (5), if X = kY (k < 0), we have the minimum cov(X,Y):

$$\begin{aligned} \frac{cov(X,Y)}{\sqrt{var(X)}\sqrt{var(Y)}} &= \frac{k * var(Y)}{(-k) * var(Y)} && (9)\\ &= -1 && (10) \end{aligned}$$

Q2: Probability and inference.
(a) If X1, . . . , Xn are independent and identically normally distributed random variables, Xi $\sim$ N(μ, σ). Estimate μ and σ via maximum likelihood.
(b) Prove the conditional version of Bayes rule:

$$P(B|A,C) = \frac{P(A|B,C)P(B|C)}{P(A|C)} \tag{11}$$

A:
(a)

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{12}$$

$$\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2} \tag{13}$$

(b)

$$P(A,B,C) = P(B|A,C) * P(A,C) = P(B|A,C) * P(A|C) * P(C) \tag{14}$$
$$= P(A|B,C) * P(B,C) = P(A|B,C) * P(B|C) * P(C) \tag{15}$$

From (14) and (15) we have:

$$P(B|A,C) * P(A|C) * P(C) = P(A|B,C) * P(B|C) * P(C) \tag{16}$$
$$P(B|A,C) = \frac{P(A|B,C) * P(B|C) * P(C)}{P(A|C) * P(C)} \tag{17}$$
$$P(B|A,C) = \frac{P(A|B,C) * P(B|C)}{P(A|C)} \tag{18}$$

Q3: Probability and conditional probability.

(a) The Internet is a wonderful source of information about symptoms of rare diseases. Are you sneezing? It could be the West Nile virus! The West Nile virus (WNV) infected approximately 2,000 people in the United States last year1. Sheldon, your hypochondriac friend, is sneezing and heard about the West Nile virus on Twitter. He demands a test for the West Nile virus, why not? The test correctly identifies the presence of WNV in 95% of cases and only gives false positives in 1/10,000 cases. Unfortunately, the test indicates came back positive for West Nile virus and Sheldon is very concerned. Assume that in the population of the United States there are 300 million people susceptible to WNV.

(i) What is the probability that Sheldon has WNV?

(ii) The WNV virus is fatal in 5% of the cases. What is the probability that Sheldon will die this year? Assume a fatality rate of any cause (car accident, etc.) of 0.1%.

(b) Alice and Bob are playing a simple dice game. Each rolls one dice and the one with higher number wins. If the numbers are the same, they roll again. If Alice just won, what is the probability that she rolled a '4'?

A:
(a)

(i)

$$P(true|positive) = 0.05956$$

(ii)

$$\begin{aligned} P(die) &= P(die\ because\ of\ WNV) + P(die\ because\ of\ other\ causes) \\ &= 0.05956 * 0.05 + 0.001 \\ &= 0.003978 \end{aligned}$$

(b)

$$\begin{aligned} P('4'|won) &= \frac{P(won|'4') * P('4')}{P(won)} \\ &= \frac{1/2 * 1/6}{15/36} \\ &= 1/5 \end{aligned}$$

Q4: (a) Plot the empirical complementary cumulative distribution (ECCDF) of comic characters appearances in comic books. The ECCDF P[X > x] is defined as the fraction of characters with more than x comic book appearances. For instance, if superman appears in 1000 comic books and there are only 10 characters with more than 1000 comic book appearances out of 2000 characters, then P [X > 1000] = 10/2000.

IMPORTANT: Your plot should be in log-log scale.

(b) Let A be the adjacency matrix connecting characters to comic books, where $A_{i,j}$ has character i appearing on comic book j. Let A^T be the transpose of matrix A.

(i) What does $W = AA^T$ represent? Give the name of the entity with the largest degree in the graph that has adjacency matrix W?

(ii) What does $U = A^TA$ represent? Give the name of the entity with the largest degree in the graph that has adjacency matrix U?

(iii) Choose the correct option: If $U = A^TA$, then (1) $U^T = AA$, (2) $U^T = A^TA$, (3) $U^T = AA^T$, or (4) $U^T = A^TA^T$.

(iv) Let $P = D^{-1}W$, where $W = AA^T$ and D is a diagonal matrix where $D_{i,i} = \sum_j W_{i,j}$. Find the eigenvector x such that $x = Px$ and $\|x\|^2 = 1$, where $\|x\|^2 = \langle x, x \rangle$ is the inner product of x with itself.

A:
(a)

see appendix.

(b)

(i) W represents how many times character i and character j appear in the same book.

Captain American has the largest degree.

(ii) U represents how many characters book i and book j share.

COC 1 has the largest degree.

(iii) (2) is correct because matrix U is symmetric.

(iv)

$$x = Px \tag{19}$$

$$Dx = Wx \tag{20}$$

$$(Dx)_i = \sum_{j=1}^{n} (W_{ij} x_j) \tag{21}$$

$$D_{i,i} x_i = \sum_{j=1}^{n} (W_{ij} x_j) \tag{22}$$

$$(\sum\nolimits_j W_{ij}) x_i = \sum_{j=1}^{n} (W_{ij} x_j) \tag{23}$$

$$x_i = x_j \ for\ all\ i\ and\ j \tag{24}$$

$$x_i = \frac{1}{\sqrt{n}},\ i = 1, 2, ..., n \tag{25}$$