CS188 HW 0

1) $y = x\sin(z)e^{-x}$

$\frac{\partial y}{\partial x} = \sin(z)\left[-xe^{-x} + e^{-x}\right]$

2) a) $y^T \cdot z = \begin{bmatrix} 1 & 3 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 3 \end{bmatrix}$

$= 1 \cdot 2 + 3 \cdot 3$

$= 11$

b) $Xy = \begin{bmatrix} 2 & 4 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 14 \\ 10 \end{bmatrix}$

c) yes $\begin{bmatrix} 3/2 & -2 \\ -1/2 & 1 \end{bmatrix}$

d) 2

3) a) $\bar{x} = 3/5 = 0.6$

b) $\sigma^2 = \frac{1}{5-1}\left[3(0.6-1)^2 + 2(0.6-0)^2\right] = 0.3$

c) [redacted] $0.5^5 = \cancel{0.3125}\ 0.03125$

d) $P = x^3(1-x)^2$

$= x^3 - 2x^4 + x^5$

$P' = 3x^2 - 8x^3 + 5x^4$

max $= 0.6$

e) $P = \frac{0.1}{0.1 + 0.15} = 0.4$

4) a) F

b) T

c) F

d) F

e) T

5) a) v

b) iv

c) ii

d) i

e) iii

6) a) mean $= p$

variance $= p(1-p)$

b) $\sigma_{2x}^2 = 4\sigma^2$

$\sigma_{x+2}^2 = \sigma^2$

7) a) i) both $\qquad \log_2 n = \frac{\ln n}{\ln 2}$

ii) $g(n) = o(f(n))$

iii) $g(n) = o(f(n))$

b) binary search

```
set left/right to ends of array
while left to the left of right
    if left == right
        return right - 1
    mid = (right + left) / 2
    if arr[mid] == 0
        left = mid + 1
    else
        right = mid - 1
```

- this algorithm works because we are eliminating half of the array with each iteration

8) a) $E[xy] = E[x]E[y]$

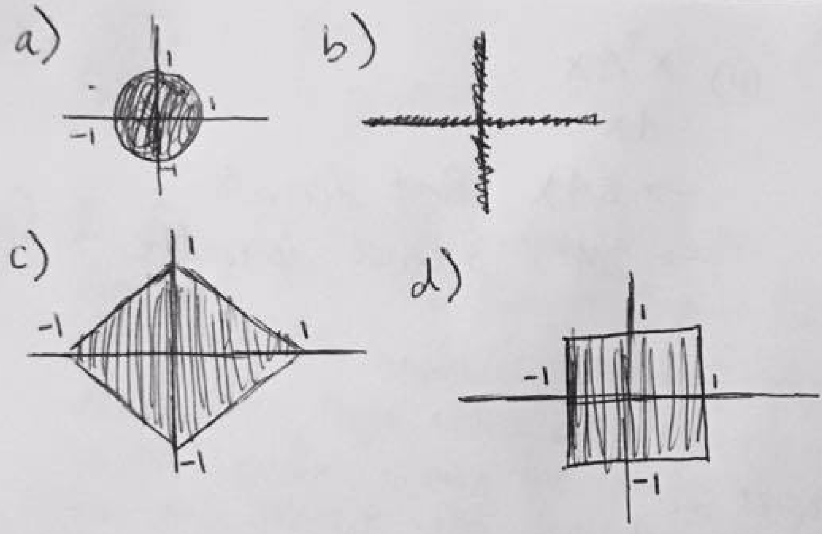
$\sum x_i y_i p_{x_i} p_{y_i} = \sum x_i p_{x_i} \sum y_i p_{y_i}$

$\sum x_i y_i p_{x_i} p_{y_i} = \sum x_i y_i p_{x_i} p_{y_i}$ ✓

b) i) The law of large numbers shows that with a large sample size, the actual behavior approaches the theoretical probability. Thus rolling 3 has a 1/6 chance which if rolled 6000 times, will appear 1000 times

ii) The Central Limit Theorem shows that as n approaches infinity, the distribution is a normal distribution

9)



b) i) eigenvalue: scaling factor for vector when multiplied by some square matrix

eigenvector: direction of this vector used when multiplied by some square matrix but vector scaled by eigenvalue

ii) $A - \lambda I = \begin{bmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{bmatrix}$

$\det(A - \lambda I) = (2-\lambda)^2 - 1$

$\lambda = 1, 3$

$\lambda = 1 \quad \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$

$x_1 + x_2 = 0$

$x_1 = -x_2$

$x = x_2 \begin{bmatrix} -1 \\ 1 \end{bmatrix}$

$\lambda = 3 \quad \left[\begin{array}{cc|c} -1 & 1 & 0 \\ 1 & -1 & 0 \end{array}\right]$

$x_1 = x_2$

$x = x_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

iii) $Ax = \lambda x$

$A^k x = A^{k-1} A x$

$= A^{k-1} \lambda x$

$= A^{k-2} A \lambda x$

$\vdots$

$= \lambda^k x$

c) (i) $\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} [x_1 \cdots x_n] = \sum a_i x_i$

$\sum a_i$

(ii) $x^T A x$

$= Ax^2$

$\rightarrow 2Ax$ first derivative

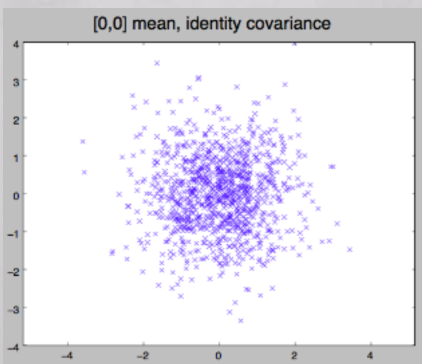
$\rightarrow 2A$ second derivative

d) (ii) $\|w\|_2 \, \|x\| + b = 0$

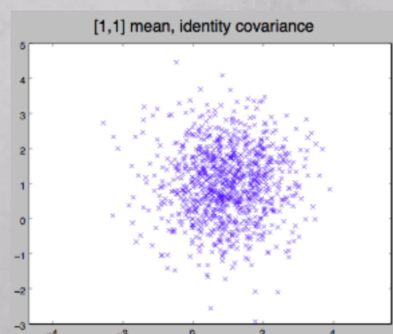
$\frac{b}{\|w\|_2} = \|x\|$

(i) The difference between 2 points on the line, $x_1 - x_2$ refers to the direction of the line, w. If the line is orthogonal to w^T, then we know that w is orthogonal to the line.

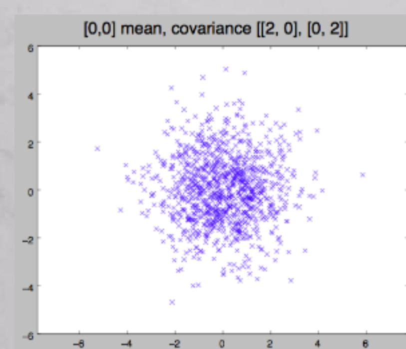
10) a)



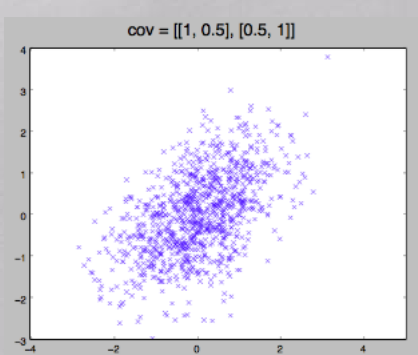
b)



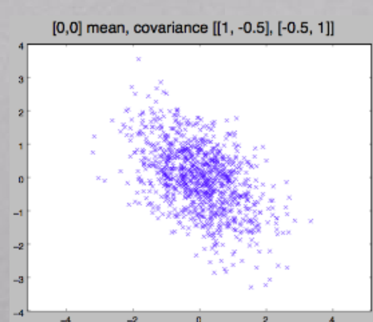
c)



d)



e)



11) $\begin{bmatrix}0\\1\end{bmatrix}$

12) I found "NYPD Complaint Data Current YTD" in the NYC open Data ~~sid~~ site. It includes complaint time, description, legal crime category, region in NY, location. Using the various features, we can predict the legal crime category from the other features. The dataset has 361,740 entries. There are a total of 24 features. Other less important features included in those 24 are complaint ID and various other keys.