CM146, Fall 2017
Problem Set 0: Math prerequisites
Due Jan 18, 2018

1 Problem 1

Solution:
$$\frac{\partial y}{\partial x} = \sin(z)\frac{\partial}{\partial x}xe^{-x} = \sin(z)(1-x)e^{-x}.$$

2 Problem 2

(a) Solution:

$$y^T z = \begin{pmatrix} 1 & 3 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 11 \end{pmatrix}.$$

(b) Solution:

$$Xy = \begin{pmatrix} 2 & 4 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 14 \\ 10 \end{pmatrix}.$$

(c) Solution:

$$X^{-1} = \frac{1}{2} \begin{pmatrix} 3 & -4 \\ -1 & 3 \end{pmatrix} = \begin{pmatrix} 3/2 & -2 \\ -1/2 & 3/2 \end{pmatrix}.$$

(d) Solution:

$$rank(X) = 2.$$

3 Problem 3

(a) Solution:

$$\frac{1+1+0+1+0}{5} = 3/5.$$

(b) Solution:

$$\frac{(1-3/5)^2 + (1-3/5)^2 + (0-3/5)^2 + (1-3/5)^2 + (0-3/5)^2}{5-1} = \frac{4+4+9+4+9}{4*25} = 0.3.$$

(c) Solution:

$$\frac{1}{2^5}.$$

(d) Solution:

$$P = p^3(1-p)^2.$$
$$0 = \frac{dP}{dp} = 3p^2(1-p)^2 - 2p^3(1-p) \Rightarrow p \in \{0, 1, 3/5\}.$$

Maximum for $p = 3/5$.

(e) Solution:

$$\frac{0.1}{0.1+0.15} = \frac{2}{5}.$$

4 Problem 4

(a) False

(b) True

(c) False

(d) False

(e) True

5 Problem 5

(a) (v)

(b) (iv)

(c) (ii)

(d) (i)

(e) (iii)

6 Problem 6

(a) Mean: p

Variance: $p(1-p)$

(b) $Var(X+2) = \sigma^2$

$Var(2X) = 4Var(X) = 4\sigma^2$

7 Problem 7

(a) Both.

$$\lim_{x\to\infty} \frac{lg(x)}{ln(x)} = \lim_{x\to\infty} 1/ln(2) < \infty.$$

$$\lim_{x\to\infty} \frac{ln(x)}{lg(x)} = \lim_{x\to\infty} 1/lg(e) < \infty.$$

(b) $g(n) = O(f(n))$ only.

$$\lim_{x\to\infty} \frac{g(x)}{f(x)} = \lim_{x\to\infty} \frac{x^{10}}{3^x} = 0 < \infty.$$

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = \lim_{x\to\infty} \frac{3^x}{x^{10}} = \infty.$$

(c) $g(n) = O(f(n))$ only.

$$\lim_{x\to\infty} \frac{g(x)}{f(x)} = \lim_{x\to\infty} \left(\frac{2}{3}\right)^n = 0 < \infty.$$

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = \lim_{x\to\infty} \left(\frac{3}{2}\right)^n = \infty.$$

(d) Algorithm:

i. Set `low` to 0 and `high` to n.

ii. Set `mid` to $\frac{\texttt{low+high}}{2}$.

iii. If `array[floor(mid)]` is 1 then set `high` to `floor(mid)`, else set `low` to `floor(mid)`.

iv. If `high-low == 1` then transition happens between `high` and `low`, else go to (ii).

Explanation: `array[low]` always remains 0 and `array[high]` always remains 1. So, once `high` $-$ `low` $==$ 1 we know the transition happens in between the two. `high` $-$ `low` will converge to 1, because `floor(mid)` $<$ `high` and `floor(mid)` $>$ `low` whenever `high`$-$`low` > 1, so the difference becomes smaller on every step.

8 Problem 8

(a) Proof for general case:

$$\begin{aligned}
\mathbb{E}[XY] &= \int_a^b \int_c^d xyp_{XY}(xy)\,dy\,dx \\
&= \int_a^b \int_c^d xyp_X(x)p_Y(y)\,dy\,dx \\
&= \int_a^b xp_X(x)\,dx \int_c^d yp_Y(y)\,dy \\
&= \mathbb{E}[X]\mathbb{E}[Y].
\end{aligned}$$

(b) Let X_i be 1 if the result of the i^{th} roll is 3 and 0 otherwise. Therefore the X_i's are independent Bernoulli random variables with $p = 1/6$, *i.e.* they are independent identically distributed random variables.

Below I used the CLT because I though it was more precise, but relized it has the same flaws as LLN. By Strong LLN $\sum X_i \to N/6$ almost surely as $N \to \infty$. Therefore, we expect 1000 3's.

Therefore, by the Central Limit Theorem, in the limit of large N

$$\frac{1}{\sqrt{6000}} \sum_{i=1}^{6000} (X_i - 1/6) \sim \mathcal{N}(0, 5/36)$$

i.e.,

$$\frac{1}{6000} \sum_{i=1}^{6000} X_i \sim \mathcal{N}\left(\frac{1}{6}, \frac{5}{36 \cdot 6000}\right)$$

$$\sum_{i=1}^{6000} X_i \sim \mathcal{N}\left(1000, \frac{5 \cdot 6000}{36}\right)$$

So, $\sigma \approx 28.86$. Therefore, there is a 95% probability that there will be between 942 and 1058 3's.

(c) Direct consequence of Central Limit Theorem, which states that in the limit of large n:

$$\sqrt{n}\left(\frac{1}{n} \sum_{i=1}^{6000} X_i - \mu\right) \sim \mathcal{N}(0, \sigma^2)$$

where $Var(X_i) = \sigma^2$ and $\mathbb{E}(X_i) = \mu$.

9 Problem 9

(a)
i. Circle centered at the origin with radius 1.
ii. The x- and y-axis.
iii. Square with diagonals aligned with the x- and y-axis and side length equal to $\sqrt{2}$.
iv. Square centered at origin, with sides parallel to the x- and y-axis and side length equal to 1.

(b)
i. λ is called an eigenvalue of a square matrix A if and only if there exists a non-zero vector $\vec{v}$ such that $A\vec{v} = \lambda\vec{v}$. All such $\vec{v}$ are called eigenvectors of A associated with λ.
ii. The eigenvalue $\lambda = 3$ has eigenvector of the form $k\begin{pmatrix}1\\1\end{pmatrix}$ for any $k \in \mathbb{R}\backslash\{0\}$.
iii. For any λ eigenvalue of A and associated eigenvector $\vec{v}$

$$\begin{aligned}
&A\vec{v} = \lambda\vec{v}\\
\Rightarrow&(A-\lambda I)\vec{v} = \vec{0}\\
\Rightarrow&(A^{k-1}I^0 + ... + A^0 I^{k-1})(A-\lambda I)\vec{v} = \vec{0}\\
\Rightarrow&(A^k - (\lambda I)^k)\vec{v} = \vec{0}\\
\Rightarrow&A^k\vec{v} = \lambda^k\vec{v}.
\end{aligned}$$

Due to commutativity of identity matrix.

(c)
i. First,

$$a^T x = \begin{pmatrix} a_1 & a_2 & \dots & a_n \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = a_1 x_1 + ... + a_n x_n.$$

So,

$$\begin{aligned}
\frac{\partial a^T x}{\partial x} &= \begin{pmatrix} \frac{\partial}{\partial x_1} & ... & \frac{\partial}{\partial x_2} \end{pmatrix} (a_1 x_1 + ... + a_n x_n)\\
&= \begin{pmatrix} a_1 & a_2 & \dots & a_n \end{pmatrix}
\end{aligned}$$

ii.

$$\begin{pmatrix} x_1 & x_2 & \dots & x_n \end{pmatrix} \begin{pmatrix} A_{11} & \dots & A_{1n} \\ \vdots & \ddots & \vdots \\ A_{n1} & \dots & A_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \sum_{i,j} A_{ij} x_i x_j.$$

$$\frac{\partial x^T A x}{\partial x} = \begin{pmatrix} A_{11}x_1 + \sum_{j \neq 1}(A_{j1} + A_{1j})x_j & \dots & A_{nn}x_n + \sum_{j \neq n} \sum_j (A_{jn} + A_{nj})x_j \end{pmatrix}$$

$$\frac{\partial^2 x^T A x}{\partial x^2} = \begin{pmatrix} A_{11} & \dots & A_{1n} \\ \vdots & \ddots & \vdots \\ A_{n1} & \dots & A_{nn} \end{pmatrix} = A.$$

(d) i.

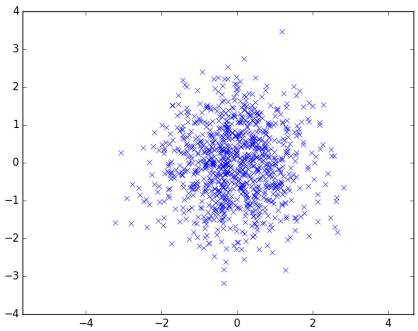
$$\begin{aligned} \omega^T(x_1 - x_2) &= \omega^T x_1 - \omega^T x_2 \\ &= (\omega^T x_1 + b) - (\omega^T x_2 + b) = 0. \end{aligned}$$

ii. The distance will be equal to the magnitude of the vector $\vec{v}$ which is orthogonal to the line, such that the vector itself belongs to the line, *i.e.*

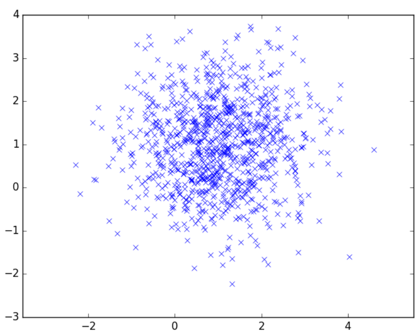
$$\vec{v} = k\omega, \quad \omega^T \vec{v} + b = 0.$$

For $k = -b^{-1}$ both equations are satisfied (uniquely since two lines intersect in at most one point) and the required magnitude is obtained.

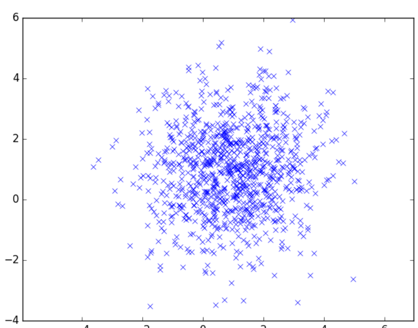
10 Problem 10



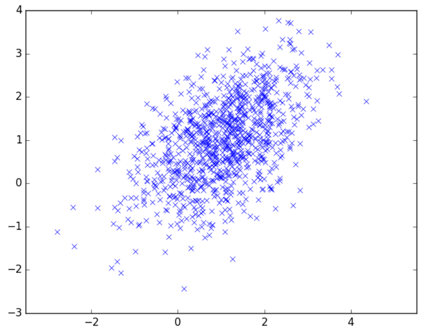
(a)



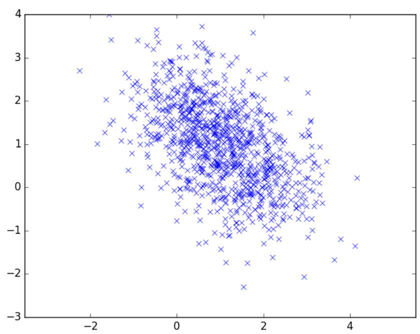
(b)



(c)



(d)



(e)

11 Problem 11

$$\binom{0}{1}$$

12 Problem 12

(a) The MNIST Database

(b) http://yann.lecun.com/exdb/mnist/

(c) The dataset contains images of handwritten digits. The content of the image defines the features for the specific datapoint, while each datapoint also has a label associated with it that specifies which digit is written in the image.

(d) One training set contains 60,000 and the other 10,000 datapoints.

(e) There is one feature for each example: the image.