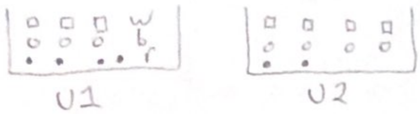
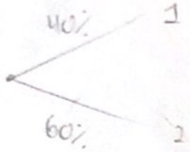
1| Two urns.





Zack Berger

a) $P(2 \text{ balls red}) = P(2R \cap \text{Urn } 1) + P(2R \cap \text{Urn } 2) = P(U1)\,P(2R \mid U1) + P(U2)(2R \mid U$

$= \left(\frac{4}{10}\right)\left(\frac{4}{10}\right)\left(\frac{3}{9}\right) + \left(\frac{6}{10}\right)\left(\frac{2}{10}\right)\left(\frac{1}{9}\right) = \frac{1}{15}$

b) $P(\text{second ball blue}) = P(U1 \cap 1^{st} \text{ ball blue} \cap 2^{nd} \text{ ball blue})$

$+ P(U1 \cap 1^{st} \text{ ball not-blue} \cap 2^{nd} \text{ ball blue})$

$+ P(U2 \cap 1 \text{ ball not-blue} \cap 2 \text{ ball blue})$

$+ P(U2 \cap 1 \text{ ball blue} \cap 2 \text{ ball blue})$

$= \left(\frac{4}{10}\right)\left(\frac{3}{10}\right)\left(\frac{2}{9}\right) + \left(\frac{4}{10}\right)\left(\frac{7}{10}\right)\left(\frac{3}{9}\right) + \left(\frac{6}{10}\right)\left(\frac{6}{10}\right)\left(\frac{4}{9}\right) + \left(\frac{6}{10}\right)\left(\frac{4}{10}\right)\left(\frac{3}{9}\right)$

$= \frac{9}{25}$

c) $P(\text{second ball blue} \mid \text{first ball red}) = \dfrac{P(\text{second ball blue} \cap \text{first ball red})}{P(\text{first ball red})}$

$= \dfrac{\left(\frac{4}{10}\right)\left(\frac{4}{10}\right)\left(\frac{3}{9}\right) + \left(\frac{6}{10}\right)\left(\frac{2}{10}\right)\left(\frac{4}{9}\right)}{\left(\frac{4}{10}\right)\left(\frac{4}{10}\right) + \left(\frac{6}{10}\right)\left(\frac{2}{10}\right)}$ by Law of Total Probability

$= 0.38095 = \frac{8}{21}$

a. $1/15$ c. $8/21$
b. $9/25$

2 | 6 dice with six sides tossed. Probability of "3 different numbers each appear twice"?

- we calculate how many different sets of numbers there are satisfying "3 different numbers each appearing twice"
  - ↳ (i.e. 112233, 114311, 662255, etc...)

  There are $\binom{6}{3} = 20$ of these sets.

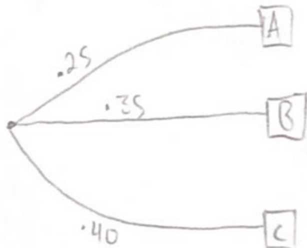
- The probability of "3 different numbers each appearing twice" is equal to the sum of the probabilities of each of the afforementioned sets
  - ↳ Probability $= P(112233) + \cdots + P(662255)$

- Claim: the probabilities of each set occuring are equivalent.
- Note that there are 6^6 permutations of the 6 dice and $\frac{6!}{(2!)^3}$ ways for a set to come up. Hence, $P(\text{set}) = \frac{6!/(2!)^3}{6^6}$

- Hence, $P(\text{"3 diff. nums each twice"}) = 20 \cdot \frac{6!/(2!)^3}{6^6} = 0.03858$

$$\boxed{P(\text{"3 diff. nums each twice"}) = 0.03858}$$

3



.05 defective

.04 defective

.02 defective

If a randomly drawn bolt is defective, what is probability it was

- manufactured by A (event A)?
- " " B (event B)?
- " " C (event C)?

- $P(A \mid \text{"Bolt defective"}) = \dfrac{P(A)P(\text{"Bolt defective"} \mid A)}{P(\text{"Bolt defective"})}$

↳ By Law of Total Probability,

$P(\text{"Bolt defective"}) = P(A)P(\text{bolt defective} \mid A) + P(B)P(\text{bolt def.} \mid B) + P(C)P(\text{bolt def.} \mid C)$

So, $P(A \mid \text{"Bolt defective"}) = \dfrac{(0.25)(0.05)}{(0.25)(0.05) + (0.35)(0.04) + (0.40)(0.02)} = \dfrac{25}{69}$

- Similarly, $P(B \mid \text{"Bolt def."}) = \dfrac{P(B)P(\text{"Bolt def."} \mid B)}{P(\text{"Bolt defective"})} = \dfrac{28}{69}$

$P(C \mid \text{"Bolt def."}) = \dfrac{P(C)P(\text{"Bolt def."} \mid C)}{P(\text{"Bolt. def"})} = \dfrac{16}{69}$

Probability bolt manufactured by

- Machine A = $25/69$
- Machine B = $28/69$
- Machine C = $16/69$

4 | Let X, Y be discrete variables.

a) Show $E[X+Y] = E[X] + E[Y]$

Recall, expectation defined as $E[u(x,y)] = \sum_{(x,y)\in S} u(x,y)\, p(x,y)$ for $(x,y)\in S$.

Then
$$E[X+Y] = \sum_{(x,y)\in S} (x+y)\, p(x,y)$$
$$= \sum_{(x,y)\in S} \left[x\, p(x,y) + y\, p(x,y) \right]$$
$$= \sum_x \sum_y x\, p(x,y) + \sum_y \sum_x y\, p(x,y) = \sum_x x \underbrace{\sum_y p(x,y)}_{= p_x(x)} + \sum_y y \underbrace{\sum_x p(x,y)}_{= p_y(y)}$$
$$= \sum_{x\in S} x\, p_x(x) + \sum_{y\in S} y\, p_y(y) = E[X] + E[Y]$$

$\square$

b) If X, Y independent, show $Var[X+Y] = Var[X] + Var[Y]$.

$$Var[X+Y] = E\left[(X+Y - E[X+Y])^2 \right]$$
$$= E[X^2 + XY - XE[X+Y] - XY + Y^2 - YE[X+Y] + XE[X+Y] - YE[X+Y] + E[X+Y]^2]$$
$$= E[X^2 + 2XY + Y^2 - 2XE[X] - 2XE[Y] - 2YE[X] - 2YE[Y]$$
$$+ E[X]^2 + 2E[X]E[Y] + 2E[Y]^2]$$
$$= E[X^2] + 2E[XY] + E[Y^2] + E[X]^2 + 2E[X]E[Y] + 2E[Y]^2$$
$$- 2E[X]^2 - 2E[Y]E[X] - 2E[X]E[Y] - 2E[Y]^2$$
$$= E[X^2] - E[X]^2 + E[Y^2] - E[Y]^2$$
$$+ 2E[XY] + 2E[X]E[Y] - 4E[Y]E[X]$$

$\hookrightarrow$ $2E[XY] = 2E[X]E[Y]$ by independence.

$$\Rightarrow Var[X+Y] = E[X^2] - E[X]^2 + E[Y^2] - E[Y]^2$$
$$= Var[X] + Var[Y]$$

$\square$

5| Waiting time until bus arrives is T. $T \sim \exp(\lambda)$ s.t. $P(T \leq t) = 1 - e^{-\lambda t}$ $\forall t \geq$

a) Already waited r seconds. Probability that bus will not arrive within d more seconds

$$P(T > r+d \mid T > r) = \frac{P(T > r+d \cap T > r)}{P(T > r)}$$

$$= \frac{P(T > r+d)}{P(T > r)} = \frac{1 - P(T \leq r+d)}{1 - P(T \leq r)}$$

$$= \frac{1 - 1 + e^{-\lambda(r+d)}}{1 - 1 + e^{-\lambda r}}$$

$$= \frac{e^{\lambda r}}{e^{\lambda(r+d)}} = e^{\lambda r - \lambda r - \lambda d} = e^{-\lambda d}$$

$$= 1 - P(T \leq d) = P(T > d)$$

□

$$\boxed{P(T > r+d \mid T > r) = P(T > d)}$$

b) Average wait time for bus?
i.e. expected value of T...

$$E[T] = \int_{-\infty}^{\infty} t\, f(t)\, dt = \int_{0}^{\infty} t \lambda e^{-\lambda t}\, dt$$

because $f(t) = \frac{d}{dt}(1 - e^{-\lambda t})$

PDF

$$= \lambda \left(\lim_{x \to \infty} \left[-\frac{1}{\lambda} e^{-\lambda t} t - \frac{1}{\lambda^2} e^{-\lambda t} \right]_0^x \right)$$

$$= \lambda \left(\lim_{x \to \infty} -\frac{1}{\lambda} \cancel{e^{-\lambda x}}^{\,0} x - \frac{1}{\lambda^2} \cancel{e^{-\lambda x}}^{\,0} + \frac{1}{\lambda} \cancel{e^{0}}(0) + \frac{1}{\lambda^2} e^{0} \right)$$

$$= \frac{\lambda}{\lambda^2}$$

$$\boxed{E[T] = \frac{1}{\lambda}}$$

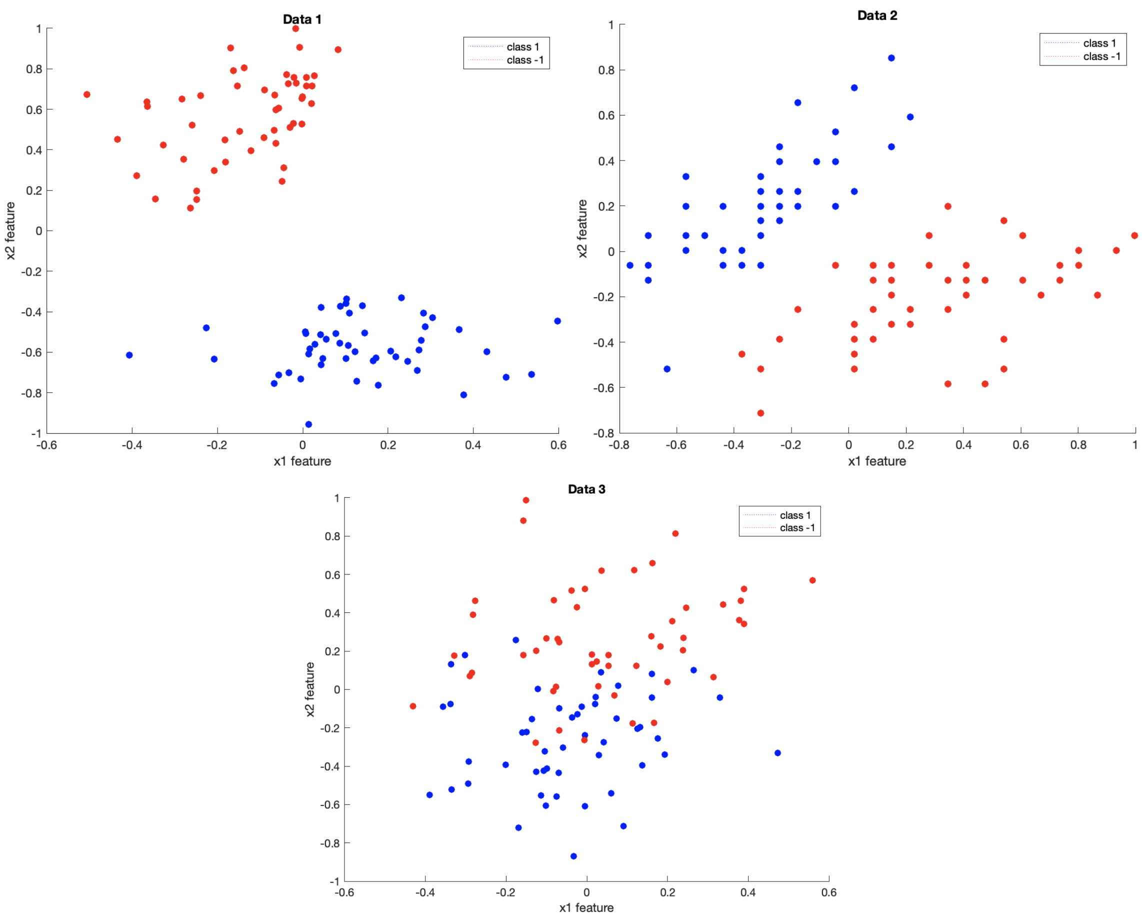
Zack Berger

Question 6

Part A

Evidently, Data 1 and Data 2 *are* linearly separable, while Data 3 is NOT linearly separable.

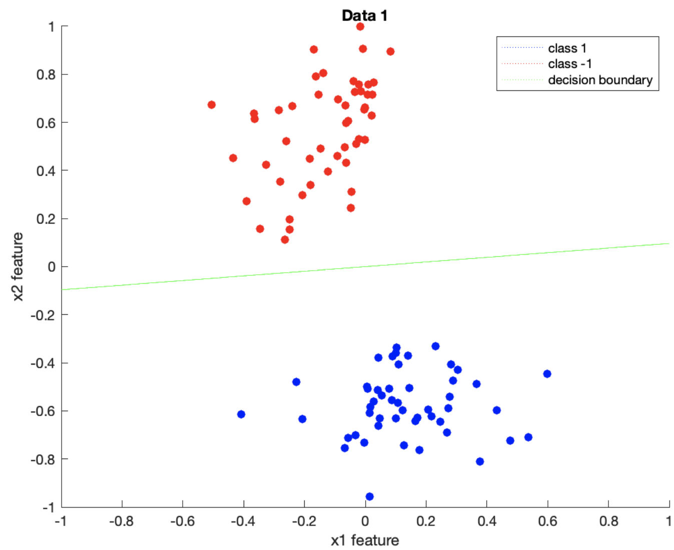
The following is a plot of each data set. The first column feature (x1) is plotted along the horizontal axis and the second column feature (x2) is plotted along the vertical axis. Points labelled as class 1 are shown in blue, and points labelled as class -1 are shown in red:



Part B

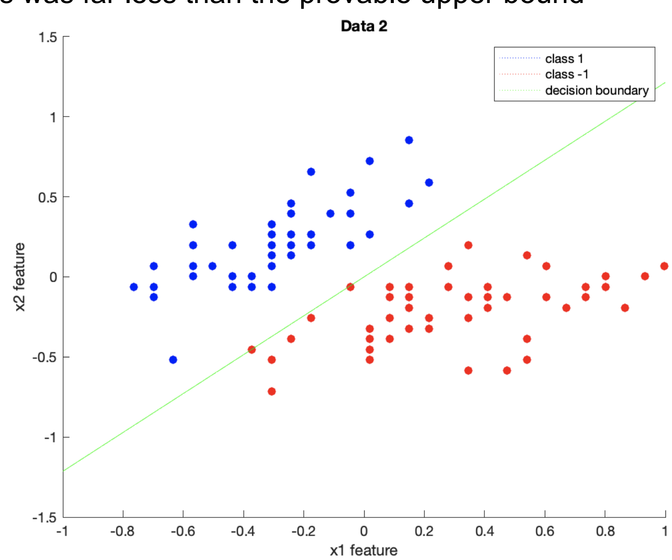
Dataset 1

- w = [0.1422, -1.4732]
- b = 0
- u = 2
    - Evidently, Perceptron converged for the provided data set as the number of updates was far less than the provable upper bound



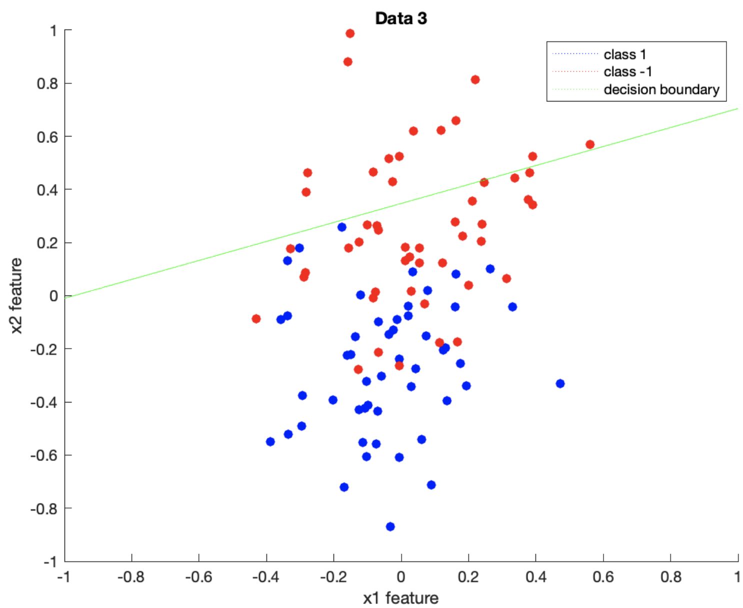
Dataset 2

- w = [-1.1092, 0.9134]
- b = 0
- u = 4
    - Evidently, Perceptron converged for the provided data set as the number of updates was far less than the provable upper bound



Dataset 3

- w = [1.0291, -2.8797]
- b = -1
- u = 4501
    - Perceptron did not converge on dataset 3 because the algorithm ran for maxIterations. The reason for its non-termination is that dataset 3 is not linearly separable



Part C

By definition (according to Daumes), $\gamma_{w,b}$ = min[y * (w•x + b)] IF w separates the data, and $\gamma_{w,b}$ = -∞ otherwise, granted that w is normalized. What follows is this calculation for each data set:

Dataset 1

- $\gamma_{w,b}$ = .1369
- The number of updates, 2, was < $1/(\gamma_{w,b}^2)$ = 53.357, as expected

Dataset 2

- $\gamma_{w,b}$ = 2.6489 x 10^{-4}
- The number of updates, 4, was < $1/(\gamma_{w,b}^2)$ = 14251772.244, as expected

Dataset 3

- $\gamma_{w,b}$ = -∞
- The data was not linearly separable!