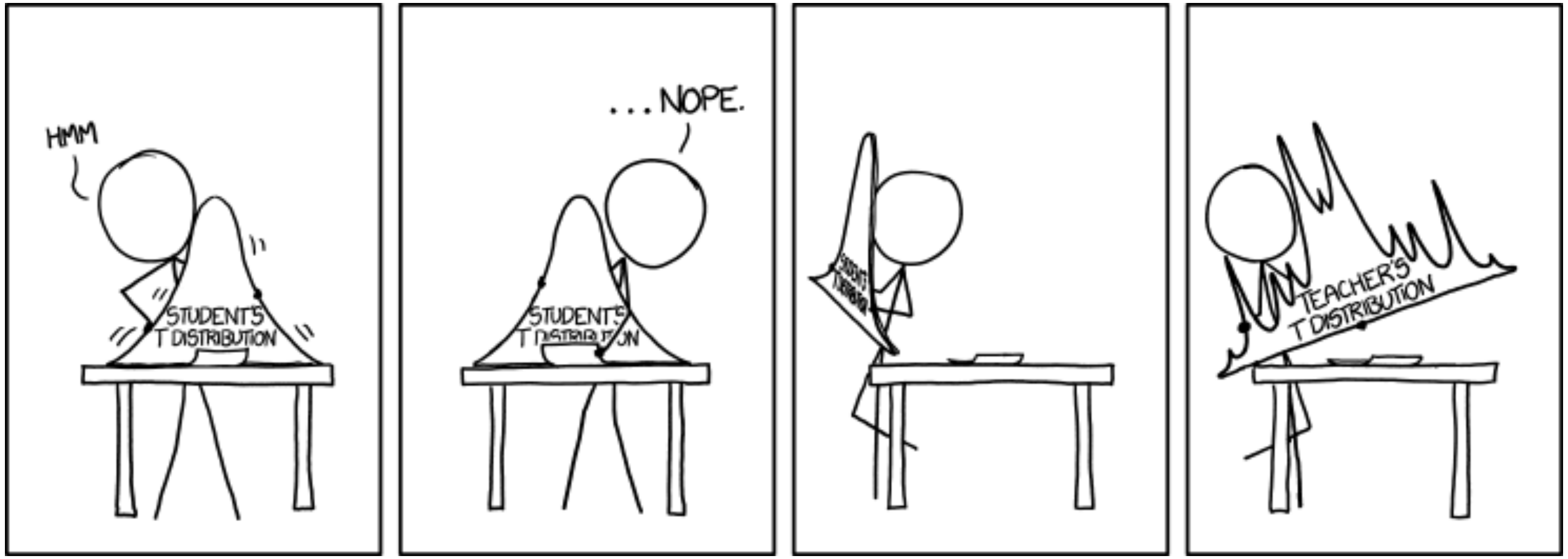
Final Review!

Info

- Thursday, December 14th from 3:30pm- 6:30pm
- Open Notes/Open Book
- No Calculators/Electronics
- Practice Final Up



Concepts to be tested

- Up until (including) last Friday
- There will be an MLE or MAP question
- There will be a sampling question
- Be ready to write pseudocode
- Be ready for probabilistic analysis of code

Chris (Spring)	Chris (Winter)	Summer
Combinatorics/Event and sample spaces	Combinatorics/Event and sample spaces	Combinatorics/Event and sample spaces
Conditional Probability	Random Variables	Code Analysis
Poisson	Poisson	Joint Probabilities and Independence
Normal	Normal/Joint	Poisson
Normal/Joint	Sampling	Conditional Probability
Sampling	Code Analysis	MLE
Samples/Random Variables/Code Analysis	MLE	Random Variables
MLE	Naive Bayes	Random Variables
Naive Bayes		Naive Bayes/Logistic Regression
Logistic Regression		

Chris (Spring)	Chris (Winter)	Summer
Combinatorics/Event and sample spaces	Combinatorics/Event and sample spaces	Combinatorics/Event and sample spaces
Conditional Probability	Random Variables	**Code Analysis**
Poisson	Poisson	Joint Probabilities and Independence
Normal	Normal/Joint	Poisson
Normal/Joint	**Sampling**	Conditional Probability
Sampling	**Code Analysis**	**MLE**
Samples/Random Variables/Code Analysis	**MLE**	Random Variables
MLE	**Naive Bayes**	Random Variables
Naive Bayes		**Naive Bayes/Logistic Regression**
MLE		

Strategies

- Write something down (get partial credit)
  - Write Random Variables!
- Figure out what concept the question is testing you on
- Don’t miss easier questions because you run out of time
- Roughly 9 questions (designed to take a TA an hour and a half)
- Points per question correspond with how many minutes you should spend
- Use notes wisely (maybe pretend we say you’re only allowed one page)
- It’s sometimes useful to make a list of concepts you’re responsible for
- We curve!

Today's Lecture!

You are experimenting with a new training course to prepare TAs for exam grading. You give the new training to 100 graders (group A) and give the old, standard training to another set of 100 graders (group B). All 200 graders are then asked to grade the same assignment.

The data collected by your experiment are the 100 grades given to the assignment by the graders in group A ($A_1 \ldots A_{100}$), and the 100 grades given by the graders in group B ($B_1 \ldots B_{100}$). You assume that each grade is IID given the grader's group.

You notice that the sample mean of the two groups is about the same. In expectation all graders are accurate. However the sample standard deviation of the grades given by group A was 5 percentage points, whereas the sample standard deviation of grades given by group B was 10 percentage points.

In this question we expect you to write pseudo-code. You will be assessed on the quality of your algorithm, not on programming syntax. Please be as precise as possible. You may use any of the following methods:

Method	**Description**
size(L)	Returns the number of elements in list L
mean(L)	Returns the arithmetic mean of the values in a list L
join(L_1, L_2)	Returns a list that has all the elements from L_1 and L_2
sum(L)	Returns the sum of all elements in L
sampleReplace(L, n)	Returns a list of n samples, drawn from list L with replacement
sampleNoReplace(L, n)	Returns a list of n samples, drawn from list L without replacement

a. Provide pseudo code for a method **sampleStandard**(S) that calculates the unbiased estimate of standard deviation for a list of IID samples $S = [S_1, S_2, \cdots, S_n]$.

Provide pseudo code for a method sampleStandard that can calculate the unbiased estimate of standard deviation for a list of IID samples S=$[S_1, S_2 \dots . S_{100}]$

Method	**Description**
size(L)	Returns the number of elements in list L
mean(L)	Returns the arithmetic mean of the values in a list L
join(L_1, L_2)	Returns a list that has all the elements from L_1 and L_2
sum(L)	Returns the sum of all elements in L
sampleReplace(L, n)	Returns a list of n samples, drawn from list L with replacement
sampleNoReplace(L, n)	Returns a list of n samples, drawn from list L without replacement

```
def sampleStandard(S):
    sampleMean = mean(S)
    diffSum = 0
    for Si in S:
        diffSum += (Si - sampleMean)^2
    return sqrt(diffSum / (101))
```

You are experimenting with a new training course to prepare TAs for exam grading. You give the new training to 100 graders (group A) and give the old, standard training to another set of 100 graders (group B). All 200 graders are then asked to grade the same assignment.

The data collected by your experiment are the 100 grades given to the assignment by the graders in group A ($A_1 \ldots A_{100}$), and the 100 grades given by the graders in group B ($B_1 \ldots B_{100}$). You assume that each grade is IID given the grader's group.

You notice that the sample mean of the two groups is about the same. In expectation all graders are accurate. However the sample standard deviation of the grades given by group A was 5 percentage points, whereas the sample standard deviation of grades given by group B was 10 percentage points.

In this question we expect you to write pseudo-code. You will be assessed on the quality of your algorithm, not on programming syntax. Please be as precise as possible. You may use any of the following methods:

Method	**Description**
size(L)	Returns the number of elements in list L
mean(L)	Returns the arithmetic mean of the values in a list L
join(L_1, L_2)	Returns a list that has all the elements from L_1 and L_2
sum(L)	Returns the sum of all elements in L
sampleReplace(L, n)	Returns a list of n samples, drawn from list L with replacement
sampleNoReplace(L, n)	Returns a list of n samples, drawn from list L without replacement

a. Provide pseudo code for a method **sampleStandard**(S) that calculates the unbiased estimate of standard deviation for a list of IID samples $S = [S_1, S_2, \cdots, S_n]$.

b. Was the difference in standard deviation significant? Write a method **pValue**(A, B) where A is the list of the grades given by group A and B is a list of the grades given by group B. Under the assumption that all 200 grades are identically distributed, calculate the probability of observing a difference in sample standard deviation greater than or equal to five. You may use **sampleStandard**(S) from part (a).

```
def pValue(A, B):
    U = join(A, B)
    count = 0
    repeat(10000):
        sampleA = sampleWithReplace(U, 100)
        sampleB = sampleWithReplace(U, 100)
        stdA = sampleStandard(sampleA)
        stdB = sampleStandard(sampleB)
        # Either one tailed or two tailed is fine!
        if |stdA - stdB| > 5:
            count++
    return count / 10000
```

In class we saw how climate sensitivity suggests that there is a fierce urgency to developing clean energy solutions. Wind energy presents many opportunities. However, wind is unpredictable and so using and expanding wind energy requires probability theory. The speed of the wind at a windfarm is a random variable that varies as a *Rayliegh Distribution*. A Rayliegh distribution is parameterized by a single scale parameter θ and has the following probability density function.

$$f_X(x) = \begin{cases} \frac{x}{\theta} e^{-x^2/2\theta} & x \geq 0 \\ 0 & else \end{cases}$$

We wish to model the wind speed on a wind farm. To this end we collect N independent measurements of wind speeds $w_1, w_2, \cdots, w_N$. Find a maximum likelihood estimate of θ if we are modeling the wind speed as coming from a Rayleigh distribution.

MLE in four easy steps!

1. Find likelihood (product of likelihood of the samples)
2. Take Log
3. Take derivative with respect to parameters
4. Either set to 0 and solve, or plug into gradient ascent

Say wind speed varies as a random variable

$W \sim Rayliegh(\theta)$

$$L(\theta) = \prod_{i=1}^{N} f_W(w_i)$$

$$LL(\theta) = \log(\prod_{i=1}^{N} f_W(w_i))$$

$$LL(\theta) = \sum_{i=1}^{N} \log(f_W(w_i))$$

$$LL(\theta) = \sum_{i=1}^{N} \log(\frac{w_i}{\theta} e^{-\frac{w_i^2}{2\theta}})$$

$$LL(\theta) = \sum_{i=1}^{N} \log(w_i) - \log(\theta) - \frac{w_i^2}{2\theta}$$

$$LL(\theta) = \sum_{i=1}^{N} \log(w_i) - N * \log(\theta) - \frac{1}{2\theta} \sum_{i=1}^{N} w_i^2$$

$$LL(\theta) = \sum_{i=1}^{N} \log(w_i) - N * \log(\theta) - \frac{1}{2\theta} \sum_{i=1}^{N} w_i^2$$

Take derivative with respect to θ

$$-\frac{N}{\theta} + \frac{1}{2\theta^2} \sum_{i=1}^{N} w_i^2$$

Set to 0

$$0 = -\frac{N}{\theta} + \frac{1}{2\theta^2} \sum_{i=1}^{N} w_i^2$$

$$\frac{N}{\theta} = \frac{1}{2\theta^2} \sum_{i=1}^{N} w_i^2$$

$$\theta = \frac{1}{2N} \sum_{i=1}^{N} w_i^2$$

(20 points) You have a newly discovered document that dates back to the 1600s and you want to identify whether or not it was written by Shakespeare. Before doing any analysis, your prior belief is that the document is equally likely to be authored by Shakespeare or not by Shakespeare.

To assist in your analysis you have *k* documents written by Shakespeare D1, ... , Dk. You also have m documents written by other authors F1, ... , Fm. In your answers you may use a function contains(X, W) which returns 1 if the document X contains the world W.

a. (10 points) Write an expression for the probability that a document contains the word "eyeball" given that it was written by Shakespeare.

Part a)

P(eyeball | Shakespeare)

Out of the documents written by Shakespeare, how many of them used the word eyeball?

$$= \frac{\sum_{i=1}^{k} contains(d_i, eyeball)}{k}$$

(20 points) You have a newly discovered document that dates back to the 1600s and you want to identify whether or not it was written by Shakespeare. Before doing any analysis, your prior belief is that the document is equally likely to be authored by Shakespeare or not by Shakespeare.

To assist in your analysis you have *k* documents written by Shakespeare D1, … , Dk. You also have m documents written by other authors F1, … , Fm. In your answers you may use a function contains(X, W) which returns 1 if the document X contains the world W.

a. (10 points) Write an expression for the probability that a document contains the word "eyeball" given that it was written by Shakespeare.

b. (10 points) The document contains n unique words W_1 , … , W_n. What is the probability that the document was written by Shakespeare given that it contains those n words? Use the Naive Bayes assumption. Assume the events that a document contains two words are conditionally independent given the author.

Part b)

$$P(Shakespeare|w_1, w_2 \dots w_n)$$

$$= \frac{P(w_1, w_2 \dots w_n|Shakespeare)P(Shakespeare)}{P(w_1, w_2, \dots w_n)}$$

$$= \frac{\mathrm{P(Shakespeare)} \prod_{i=1}^{n} P(w_i|Shakespeare)}{P(w_1, w_2, \dots w_n)}$$

$$= \frac{.5 * \prod_{j=1}^{n} \frac{\sum_{i=1}^{k} contains(d_i, w_j)}{k}}{P(w_1, w_2, \dots w_n)}$$

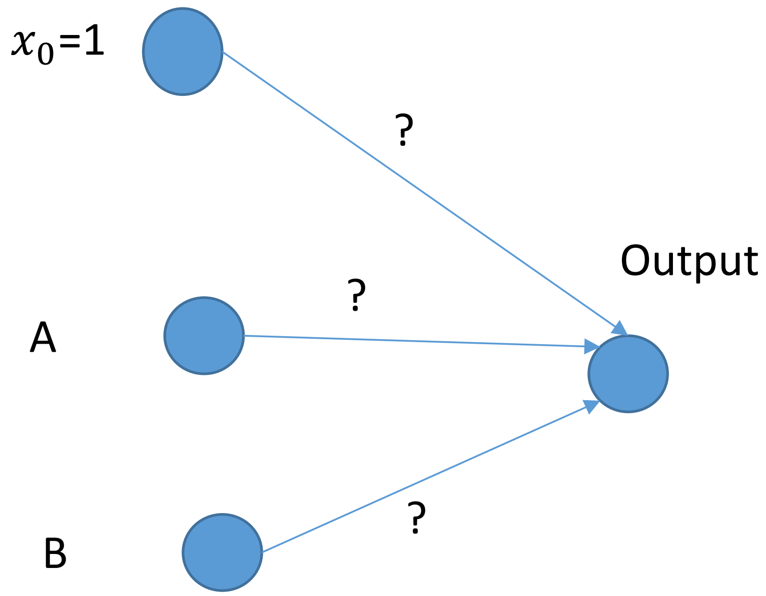
Part b)

$$=\frac{.5*\prod_{j=1}^{n}\frac{\sum_{i=1}^{k} contains\left(d_i,w_j\right)}{k}}{P(w_1,w_2,\ldots w_n)}$$

$$=\frac{.5*\prod_{j=1}^{n}\frac{\sum_{i=1}^{k} contains\left(d_i,w_j\right)}{k}}{P\left(words|Shakespeare\right)P(Shakespeare)+P\left(words|Shakespeare^{c}\right)P(Shakespeare^{c})}$$

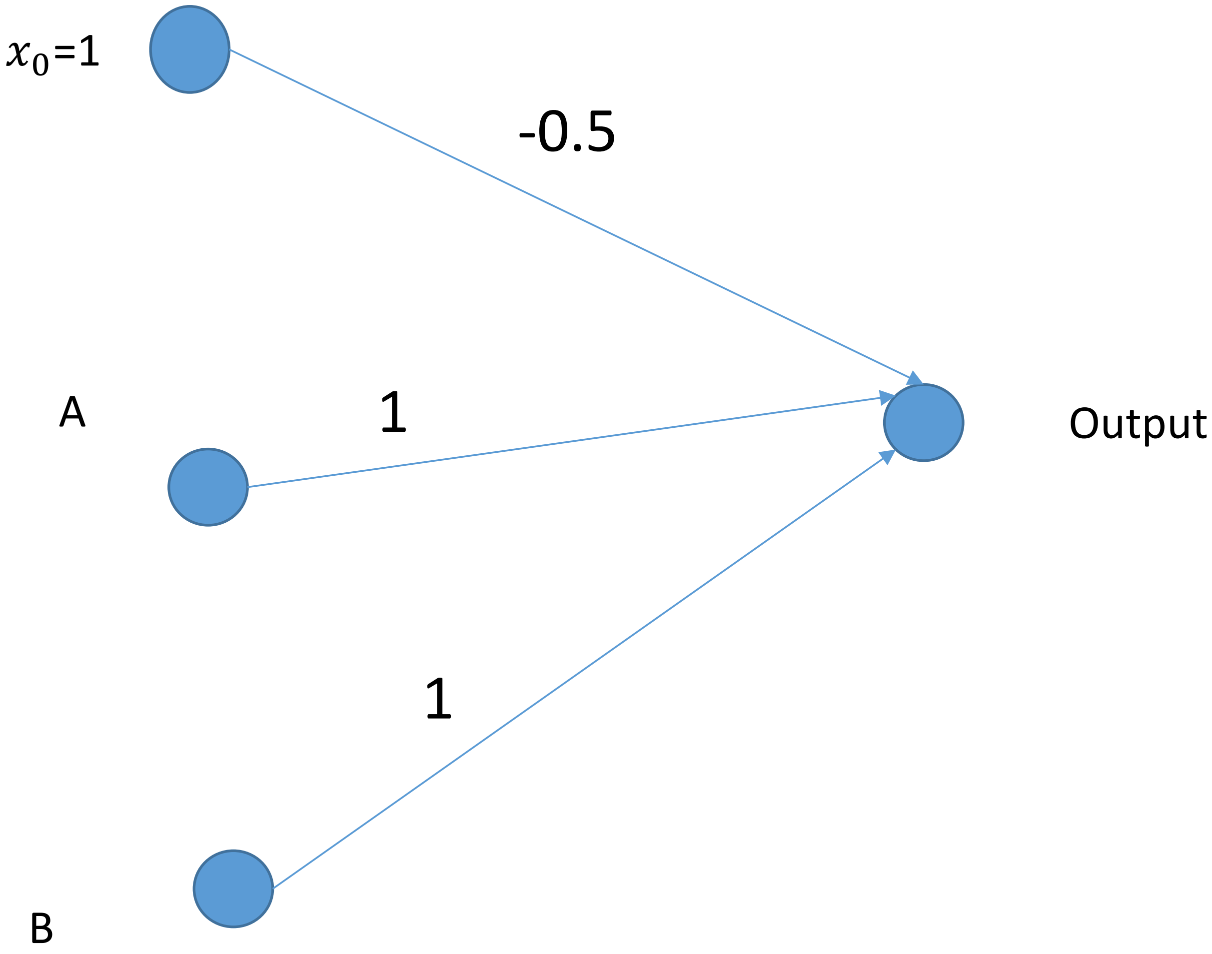
$$=\frac{.5*\prod_{j=1}^{n}\frac{\sum_{i=1}^{k} contains\left(d_i,w_j\right)}{k}}{.5*\prod_{j=1}^{n}\frac{\sum_{i=1}^{k} contains\left(d_i,w_j\right)}{k}+.5*\prod_{j=1}^{n}\frac{\sum_{i=1}^{m} contains\left(f_i,w_j\right)}{k}}$$

Neural Nets!

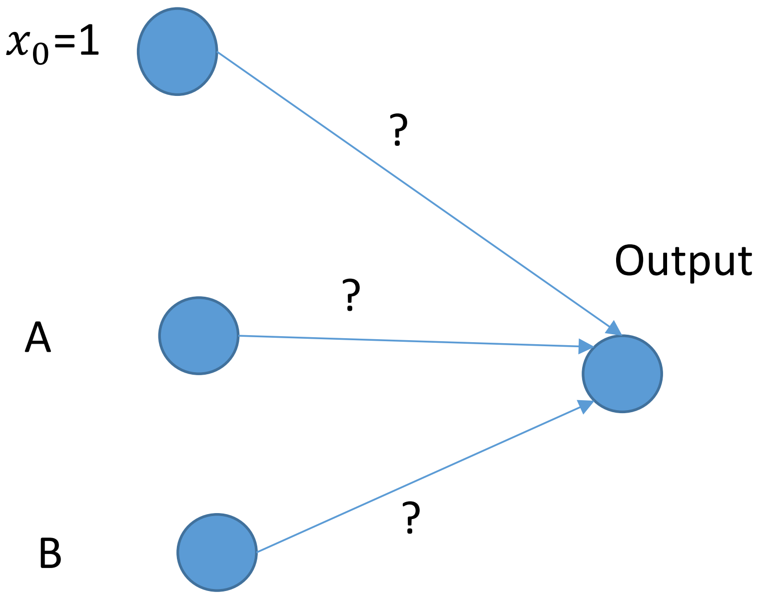


What weights might this logistic regression model learn that would allow it to perfectly classify data of the form (A or B)?

Neural Nets! OR

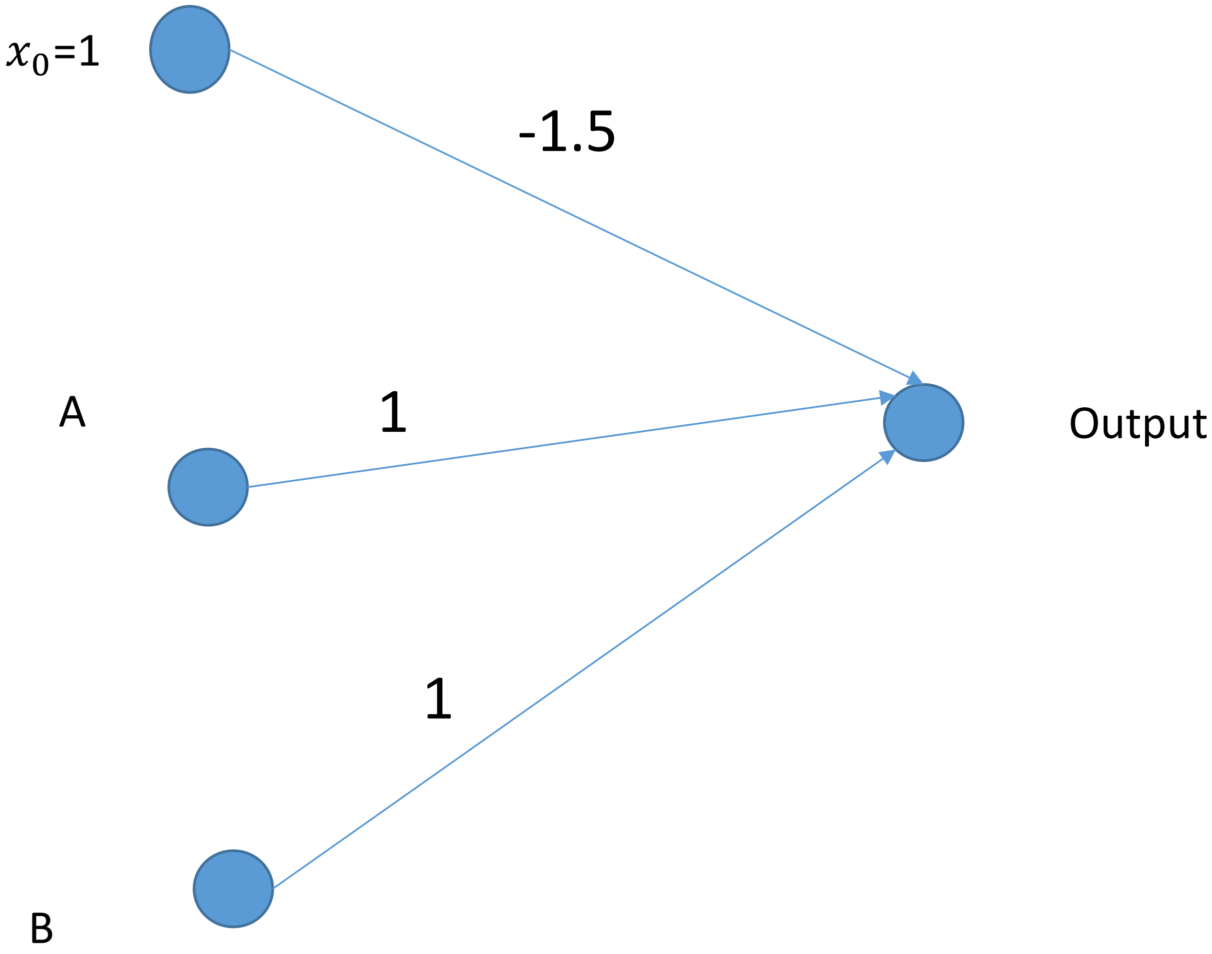


Neural Nets!



What weights might this logistic regression model learn that would allow it to perfectly classify data of the form (A and B)?

Neural Nets! AND



What’s on Wednesday