Problem 1. This exercise traces through the first few steps of a perceptron training algorithm. The task is to classify a sentence into one of two classes, which are called +1 and −1. We will use just two features; unlike the example in class, these features are not binary, but integer-valued. The features are:

pron The number of personal pronouns in the sentence.

noun The number of proper and common nouns in the sentence.

In the data below, each instance of **pron** is marked in **red boldface**, and each instance of ***noun*** is marked in ***green bold italics***. A hyphenated term such as *Cochin-China* or *great-aunt* is considered a single term. The data (classes and sentences) are taken from Argamon et al.: Gender, genre, and writing style in formal written texts, *Text* **23**(3): 321–346, 2003.

a. (15 points) Count the features in each sentence and update the perceptron weights, using Algorithm 5 of the perceptron reading (Daumé). weight starts with 0 and progresses gradually

+1 ***Clara*** never failed to be astonished by the extraordinary ***felicity*** of **her** own ***name***.

Feature counts: **pron** 1 ***noun*** 3 Weights: **pron** 1 ***noun*** 3 bias 1

−1 By 1925 present-day ***Vietnam*** was divided into three ***parts*** under French colonial ***rule***.

Feature counts: **pron** 0 ***noun*** 3 Weights: **pron** 1 ***noun*** 0 bias 0

+1 **She** found it hard to trust **herself** to the mercy of ***fate***, which had managed over the ***years*** to convert **her** greatest ***shame*** into one of **her** greatest ***assets***, and even after ***years*** of comparative ***security*** **she** was still prepared for, still half expecting the old ***gibes*** to be revived.

Feature counts: **pron** 5 ***noun*** 7 Weights: **pron** 1 ***noun*** 0 bias 0

−1 The southern ***region*** embracing ***Saigon*** and the ***Mekong delta*** was the ***colony*** of ***Cochin-China***; the central ***area*** with its imperial ***capital*** at ***Hue*** was the ***protectorate*** of ***Annam***; and the northern ***region***, ***Tongking***, was also a separate ***protectorate*** with its ***capital*** at ***Hanoi***.

Feature counts: **pron** 0 ***noun*** 16 Weights: **pron** 1 ***noun*** −16 bias −1

+1 But whenever **she** was introduced, nothing greeted the amazing, all-revealing ***Clara*** but ***cries*** of "How delightful, how charming, how unusual, how fortunate," and **she** could foresee a ***time*** when ***friends*** would name **their** ***babies*** after **her** and refer back to **her** with ***pride*** as the ***original*** from which ***inspiration*** had first been drawn.

Feature counts: **pron** 5 ***noun*** 8 Weights: **pron** 6 ***noun*** −8 bias 0

−1 The Annamese ***emperor***, ***Khai Dinh***, in ***theory*** ruled the two northern ***regions*** from ***Hue*** with the ***benefit*** of French ***protection***, while ***Cochin-China*** was governed directly from ***Paris*** but in ***effect*** all three ***territories*** were ruled as ***colonies***.

Feature counts: **pron** 0 ***noun*** 12 Weights: **pron** 6 ***noun*** −8 bias 0

6 − 26/7 = 2.29

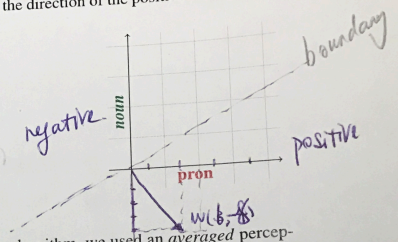
−8 + 26/7 = −4.14

b. (6 points) What is the decision boundary found by the perceptron? Give a formula, and draw the boundary on the graph with a vector pointing in the direction of the positive class (similar to Figures 4.6 and 4.9 in the reading).

Formula: $\beta = \{x : \sum_d w_d x_d = 0\} + b$

$w = (6, -8)$

Decision boundary as the plane on the right



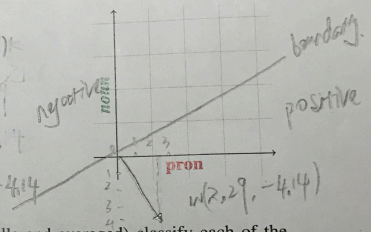
c. (5 points) Suppose instead of the vanilla perceptron algorithm, we used an *averaged* perceptron (section 4.6 in the reading). What would be the decision boundary? Give a formula and draw the boundary on the graph.

Formula: $\hat{y} = \text{Sign}\left(\sum_{k=1}^{K} c^{(k)}(w^{(k)} \cdot \vec{x} + b^{(k)})\right)$

$w = \vec{w} - \frac{1}{c}\vec{\mu}$

$b = \vec{b} - \frac{1}{c}\vec{\beta}$

$w_{pron} = \frac{16}{7} = 2.29$ $w_{noun} = -4.14$



d. (4 points) How would each of the perceptrons (vanilla and averaged) classify each of the following texts?

pron: 7
noun: 15
pron: 7

Finally **her** ***confidence*** grew to such an ***extent*** that **she** was able to explain that **she** had been christened not in the ***vanguard*** but in the extreme ***rearguard*** of ***fashion***, after a Wesleyan ***great-aunt***, and that **her** ***mother*** had formed the ***notion*** not as an unusual and charming ***conceit*** but as a preconceived ***penance*** for **her** ***daughter***, whose only ***offences*** at that tender ***age*** were **her** ***existence*** and **her** ***sex***.

Vanilla: +1 Averaged: +1

noun: 2
noun: 9

Some backward ***tribes*** inhabited the remoter ***mountains*** and ***jungles*** but the main ***population*** was of the same ***race***; today **they** are known as ***Vietnamese*** but then the outside ***world*** knew **them** as ***Annamites*** or ***Annamese***.

Vanilla: −1 Averaged: −1

problem 2

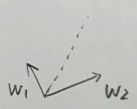
a>

Based on the Naive Bayes Model mentioned above, we should classify the text based on which one has higher probability.

if $n \cdot w_1 > n \cdot w_2$, we classify text as Class 1

else if $n \cdot w_2 > n \cdot w_1$, we classify text as Class 2.

b>

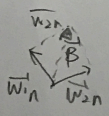


Based on a>, we try to compute the boundary that equally split the angle between w_1 and w_2.

We compute the unit vector for both w_1 and w_2 as

$$\vec{w}_{1n} = \frac{\vec{w}_1}{|\vec{w}_1|}$$

$$\vec{w}_{2n} = \frac{\vec{w}_2}{|\vec{w}_2|}$$



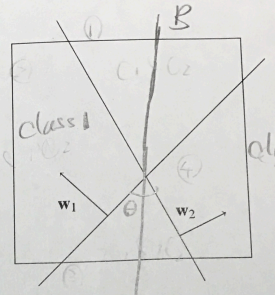
Therefore, the boundary vector can be

$$\vec{B} = \frac{\vec{w}_1}{|\vec{w}_1|} + \frac{\vec{w}_2}{|\vec{w}_2|}$$

Or $$\vec{B} = -\frac{\vec{w}_1}{|\vec{w}_1|} - \frac{\vec{w}_2}{|\vec{w}_2|}$$

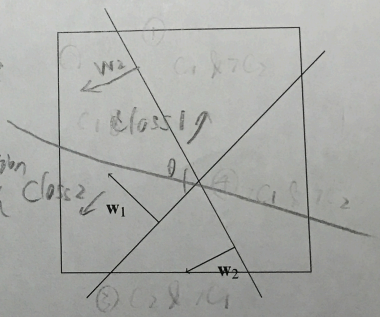
c. (4 points) The following box represents a 2-dimensional feature space, with the planes and weight vectors associated with C_1 and C_2. Use a diagram and an explanatory sentence to show how these planes determine a decision boundary, and indicate the decision regions (that is, which part of the feature space will be classified as C_1 and which as C_2).

As b state, we need to find the boundary that equally split the angle generated by $\vec{w_1}$ and $\vec{w_2}$. (B).
B's left side will be class 1 and B's right side will be class 2

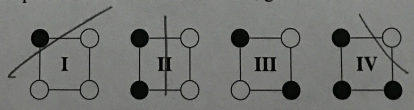


d. (4 points) Do the same for the following case (diagram and explanatory sentence). What is the difference?

Similar to c, we find the line that equally splits the vector $\vec{w_1}$ and $\vec{w_2}$.
The difference is since the direction of w_2 is opposite in compared with c, the boundary changes.



e. (4 points) The following diagrams represent possible ways to split the four possible observations of two binary features between two classes. Which of the cases below are consistent with conditional independence of the feature values, given the class?



1 2 4. 3 doesn't work because if conditional independent, it should be able to be split with one line (linear)
3 doesn't work