Problem 1

domain objects: the words and phrases of the language.
training data: whatever the person wishes to use as learning material, whether it be textbooks, online courses or tv shows.
model: the mental model of the person on how they understand the language
learning algorithms: the methods in which the person uses to learn the language
output: the set of phrases and words that the person now understands
type of learning: supervised learning

Problem 2

a. Regression based machine learning problem
b. This is a predictive task
c. A geometric model since our output is mostly real numbers
d. Grading model
e. R^2 since we have 2 types of blood pressure levels
f. R^2 since we have 2 different blood pressure numbers

Problem 3

let the feature vectors $x1, x2$ be $x1_i[33.6, 30.6, 4.8, 6.8, 1.22, 2.11, 3.00]$ and $x2_i = [36.7, 27.0, 4.7, 11.3, 1.0, 1.67, 3.83]$

a. Manhattan Distance

using the Minkowski distance formula, since Manhattan and Euclidian is just Minkowski with $p = 1, 2$ respectively.

$$d(x, y) = \left(\sum_{i=1}^{d} |x_i - y_i|^p\right)^{\frac{1}{p}}$$

I created python code to make it easier to calculate the values:

```
x = [33.6, 30.6, 4.8, 6.8, 1.22, 2.11, 3.0]
y = [36.7, 27.0, 4.7, 11.3, 1.0, 1.67, 3.83]
pList = [1, 2, 10, 100]

def MinkowskiDistance(x, y, p):
    if len(x) != len(y):
        print("x and y vectors are not equal length")

    sum = 0
    for i in range(0, len(x)):
        sum  = sum + (abs(x[i] - y[i]) ** p)

    sum = sum ** (1/p)
    print("Distance L", p, ":", sum)

def main():
    for p in pList:
        MinkowskiDistance(x, y, p)
```

a. $L1$ distance $= 12.79$
b. $L2$ distance $= 6.615$
c. $L10$ distance $= 4.556$
d. $L100$ distance $= 4.5$

e. after adding the constant vector the new vectors were
$x1 = [38.6, 35.6, 6.8, 8.8, 1.72, 2.21, 4.0]$
$x2 = [41.7, 32.0, 6.7, 13.3, 1.5, 1.77, 4.83]$

the distance values remained the same with
$L1$ distance $= 12.79$
$L2$ distance $= 6.615$
L_{10} distance $= 4.556$
L_{100} distance $= 4.5$

When using the constant $k = 2$ the new vectors will be
$x1 = [67.2, 61.2, 9.6, 13.6, 2.44, 4.22, 6.0]$
$x2 = [73.4, 54.0, 9.4, 22.6, 2.0, 3.34, 7.66]$
All the distances end up changing
$L1$ distance $= 25.58$
$L2$ distance $= 13.23$
L_{10} distance $= 9.1126$
L_{100} distance $= 9.0$

Problem 4

a. $\mathrm{P}(\mathit{grade}|\mathit{class},\ \mathit{effort})$
Class: 165B **Effort**: small
A $= 0$, B $= \frac{1}{6}$, C $= \frac{1}{6}$, D $= \frac{1}{3}$, F $= \frac{1}{3}$
Class: 165B **Effort**: medium
A $= \frac{5}{29}$, B $= \frac{10}{29}$, C $= \frac{10}{29}$, D $= \frac{4}{29}$, F $= 0$
Class: 165B **Effort**: large
A $= \frac{20}{41}$, B $= \frac{15}{41}$, C $= \frac{5}{41}$, D $= \frac{1}{41}$, F $= 0$

Class: Basketweaving **Effort**: small
A $= \frac{1}{3}$, B $= \frac{1}{3}$, C $= \frac{1}{3}$, D $= 0$, F $= 0$
Class: Basketweaving **Effort**: medium
A $= \frac{4}{7}$, B $= \frac{2}{7}$, C $= \frac{1}{7}$, D $= 0$, F $= 0$
Class: Basketweaving **Effort**: large
A $= \frac{6}{7}$, B $= \frac{1}{7}$, C $= 0$, D $= 0$, F $= 0$

b.

	Small	Medium	Large
A	50	125	250
B	75	100	100
C	75	75	25
D	50	20	5
F	50	0	0

Effort: Small
A $= \frac{10}{200}$, B $= \frac{15}{200}$, C $= \frac{15}{200}$, D $= \frac{10}{200}$, F $= \frac{10}{200}$
Effort: Medium
A $= \frac{25}{200}$, B $= \frac{20}{200}$, C $= \frac{15}{200}$, D $= \frac{4}{200}$, F $= 0$
Effort: Large
A $= \frac{50}{200}$, B $= \frac{20}{200}$, C $= \frac{5}{200}$, D $= \frac{1}{200}$, F $= 0$

c.
Using the table from part **b.**, sum the 3 different efforts and divide them by the total
Small $= \frac{300}{1000} \to 0.3$, Medium $= \frac{320}{1000} \to 0.32$, Large $= \frac{380}{1000} \to 0.38$

d.
P(A|165B) = P(A and 165B)/P(165B) $= \frac{1}{4}$
P(A|bascketweaving) = P(A and bascketweaving)/P(bascketweaving) $= \frac{3}{5}$

Problem 5

a.

	Labeled Spam	Labeled Non-spam
Detected as Spam	1750(TP)	250(FN)
Detected as Non-spam	250(FP)	7750(TN)

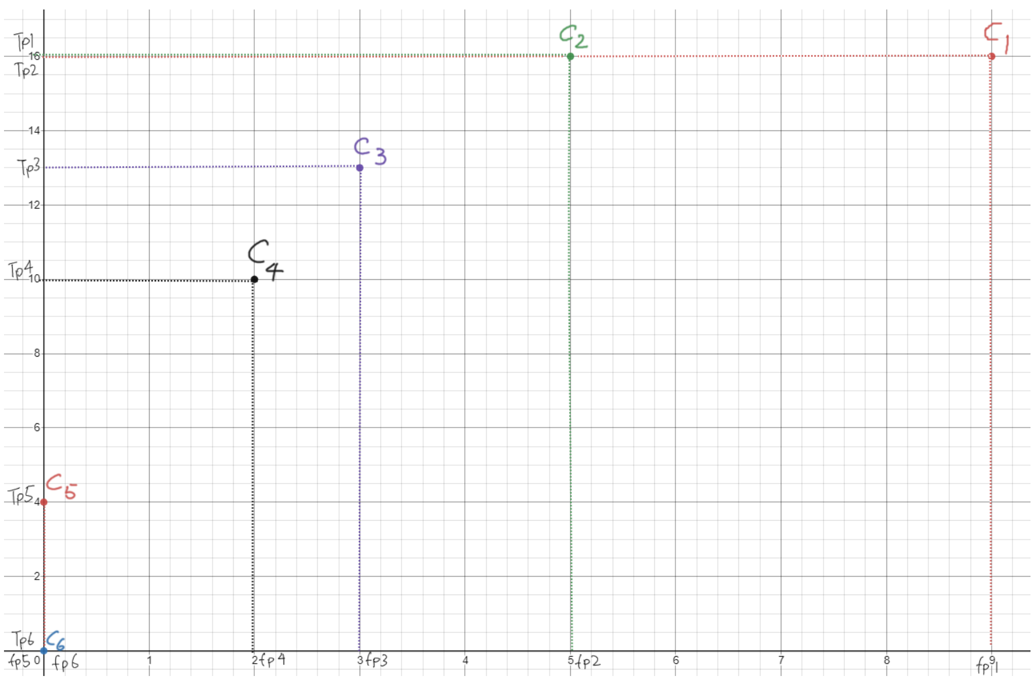
b. false positive rate $= \frac{FP}{N} = \frac{250}{8000} = \frac{1}{32}$
c. false negative rate $= \frac{FN}{P} = \frac{250}{2000} = \frac{1}{8}$
d. error rate $= \frac{FP+FN}{P+N} = \frac{250+250}{2000+8000} = \frac{500}{10000} = \frac{1}{20}$
e. precision $= \frac{TP}{TP+FP} = \frac{1750}{1750+250} = \frac{1750}{2000} = \frac{7}{8}$
f. accuracy = 1 - precision $= 1 - \frac{1}{20} = \frac{19}{20}$

Problem 6

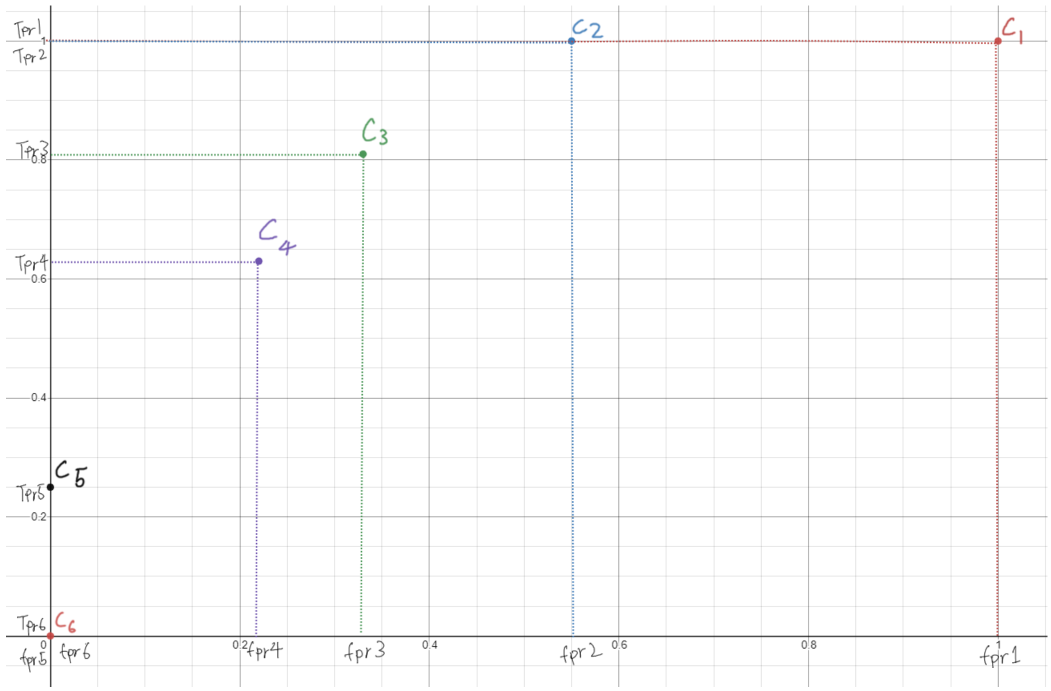
a. 9 ranking errors
b. ranking error $= \frac{9}{12 \cdot 13} = 0.058$
c. ranking accuracy = 1 - ranking error $= 1 - 0.058 = 0.942$

Problem 7

a.



b.



c.

Highest: C2 $= \frac{20}{25}$ **Lowest:** C6 $= \frac{9}{25}$

d.

Highest: C5 $= 1$ **Lowest:** C6 $= 0$

e.

Highest: C1 and C2 **Lowest:** C6 $= 0$

f.

C1 and C2

g.

C5 and C6