Tutorial: Introduction to Machine Learning and Data Mining

- Yunqiu Xu

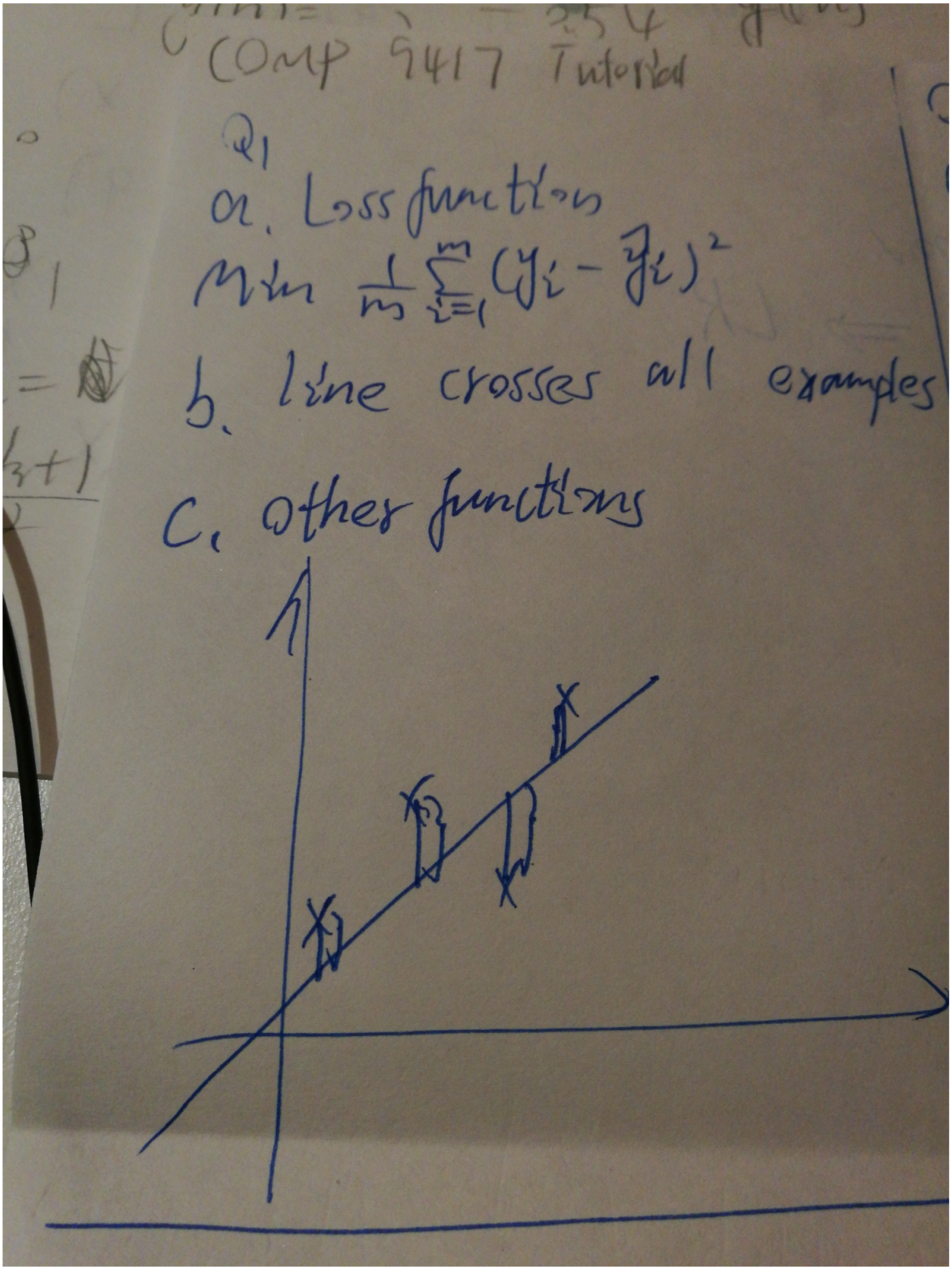
Q1

Question 1

a) What is the function that Linear Regression is trying to minimize ?

b) Under what conditions would the value of this function be zero ?

c) Can you suggest any other properties of this function ?


COMP 9417 Tutorial
Q1
a. Loss functions
$\min \frac{1}{m} \sum_{i=1}^{m} (y_i - \hat{y}_i)^2$
b. line crosses all examples
c. other functions

Q2

Question 2 Machine learning has a fair amount of terminology which it is important to get to know.

a) Why do we need features ?

b) What is a the difference between a "task", a "model" and a "learning problem" ?

c) Can different learning algorithms be applied to the same tasks and features ?

Tasks are addressed by models, whereas learning problems are solved by learning algorithms that produce models.

Machine learning is concerned with using the right features to build the right models that achieve the right tasks.

Models lend the machine learning field diversity, but tasks and features give it unity.

Does the algorithm require all training data to be present before the start of learning ? If yes, then it is categorised as **batch learning** algorithm.

If however, it can continue to learn a new data arrives, it is an **online learning** algorithm.

If the model has a fixed number of parameters it is categorised as **parametric**.

Otherwise, if the number of parameters grows with the amount of training data it is categorised as **non-parametric**.

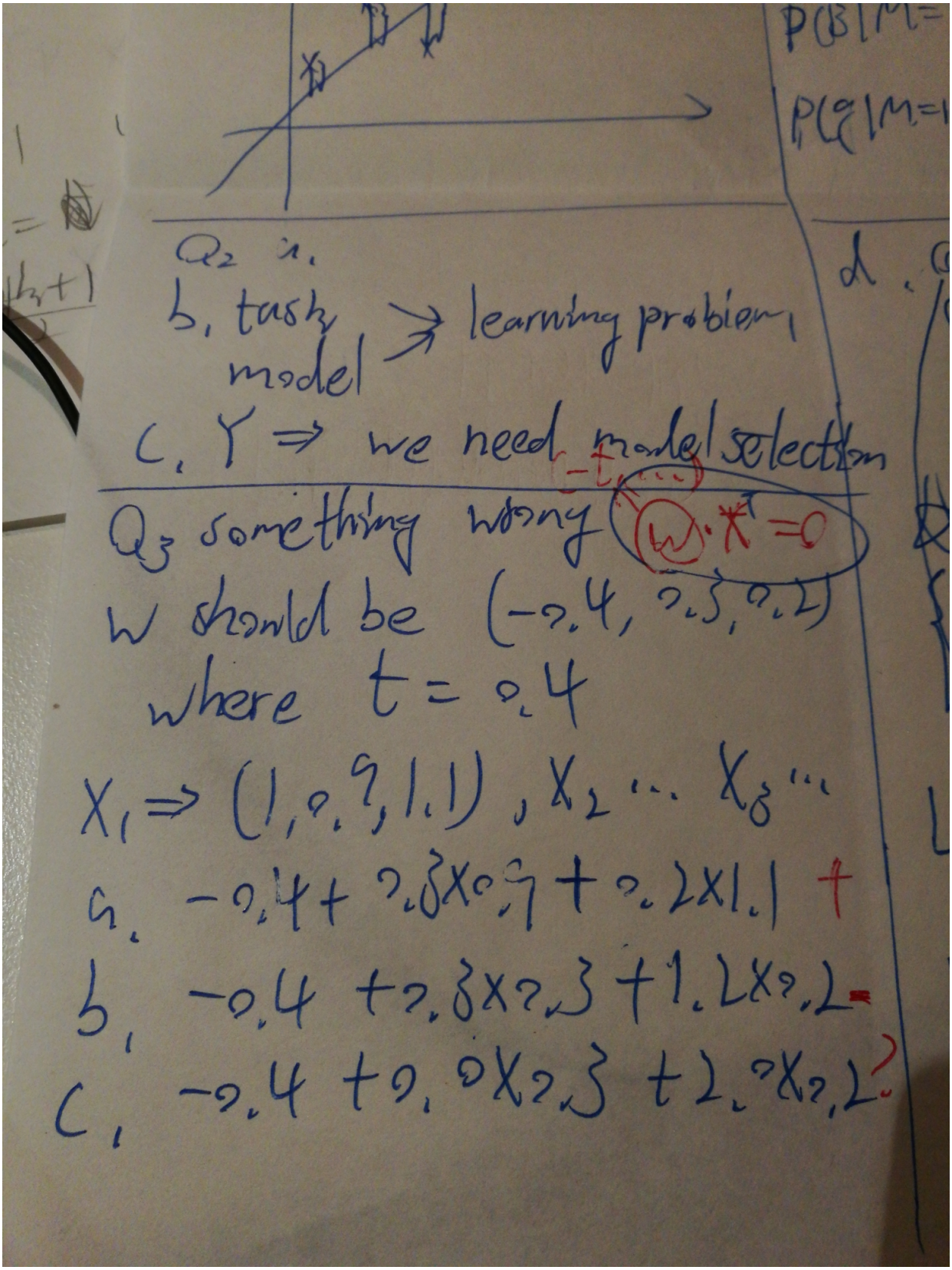
Q3

Question 3 Suppose you run a learning algorithm that returns a basic linear classifier using the *homogeneous coordinates* representation on a set of training data and obtain the follow weight vector $w = (0.4, 0.3, 0.2)$. For each of the following examples, what would the classification be using the weight vector w ?

a) $x_1 = (0.9, 1.1)$?

b) $x_2 = (0.3, 1.2)$?

c) $x_3 = (0.0, 2.0)$?


Q2 i.
b. task → learning problem
model
c. Y ⇒ we need model selection
Q3 something wrong
w should be (-0.4, 0.5, 0.2)
where t = 0.4
X1 ⇒ (1, 0.9, 1.1), X2 ... X3 ...
a. -0.4 + 0.5X0.9 + 0.2X1.1 +
b. -0.4 + 0.5X0.3 + 1.2X0.2 =
c. -0.4 + 0.5X0.3 + 2.2X0.2 ?

Q4

Table 1: Posterior probability distribution of classes given word occurrence (bold font indicates more probable class).

valuation	manufacturing	$P(Y = \text{business} \mid \text{valuation}, \text{manufacturing})$	$P(Y = \text{general} \mid \text{valuation}, \text{manufacturing})$
0	0	0.3	**0.7**
0	1	0.5	0.5
1	0	**0.6**	0.4
1	1	0.9	**0.1**

Table 2: Marginal likelihoods: think of these as probabilities of observing the data items (words) independently of any others, given the repective classes.

Y	$P(\text{valuation} = 1 \mid Y)$	$P(\text{valuation} = 0 \mid Y)$
business	0.3	0.7
general	0.1	0.9

Y	$P(\text{manufacturing} = 1 \mid Y)$	$P(\text{manufacturing} = 0 \mid Y)$
business	0.4	0.6
general	0.2	0.8

a) using the data from Table 1, what two patterns of occurrence of keywords in a text file lead to a prediction of 'business' ?

b) what prediction should be made if we have an occurrence of 'manufacturing' but NOT 'valuation' in a text file ?

c) suppose we are given a text file to classify, and we know that 'manufacturing' occurs in the text file, but we know some words are missing from the file for some reason, and we are uncertain if 'valuation' occurred or not. However, we do know that the probability of 'valuation' occuring in any text file is 0.05. Compute the probability of each class for the given text file.

d) using the values from Table 2 compute the likelihood ratios for each of the four possible patterns of occurrence the keywords

Q4

a. $V=1, m=0 \Rightarrow 0.6$

$V=1 \quad m=1 \Rightarrow 0.9$

b. $V=0, m=1$

$P(B \mid V=0, m=1) = 0.5$

$P(g \mid V=0, m=1) = 0.5$

c.

$$P(B \mid m=1) = \begin{cases} P(m=1) = 0.5 \\ P(V=1) = 0.5 \\ P(B \mid V=0, m=1) \cdot P(V=0) \\ + P(B \mid V=1, m=1) \cdot P(V=1) \end{cases}$$

$$P(g \mid m=1) = P(g \mid V=0, m=1) P(V=0) + P(g \mid V=1, m=1) P(V=1)$$

d.

① $V=1 \quad m=1$

② $V=1 \quad m=0$

③ $V=0 \quad m=1$

④ $V=0 \quad m=0$

You need to know likelihood ratio posterior odds

$$LR = \frac{P(V=1 \mid b)}{P(V=1 \mid g)} \times \frac{P(m=1 \mid b)}{P(m=1 \mid g)}$$

$$PO = \frac{P(b \mid V=1, m=1)}{P(g \mid V=1, m=1)}$$