Name: Weiqi Ji BUID: U65196797

Q1:

a) Artificial Intelligence is not always trained to be fair. sometimes AI's may produce biased results because of variations in the dataset.

b) For example you want to learn how drivers behave on the road around the world, but you only have data from one country, so you have to adjust it for that country.

c) This is when GPU technology caught up which allowed people to compute deep learning natral networks better.

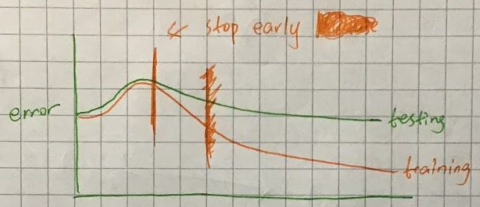
d) This happens when you split data into a training set and a validation set, and you train your model using that, but then adding some from the training set to the validation set to test again.

e) Dropout is a data augmentation approach because it helps to reduce overfitting in the data by randomly dropping out pieces of data.

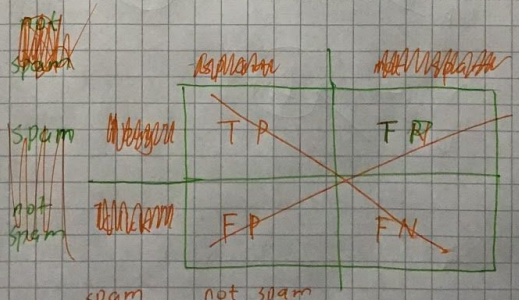
Q2:

a) A: 10^2 B: 10^0 C: 10^{-2}

b)



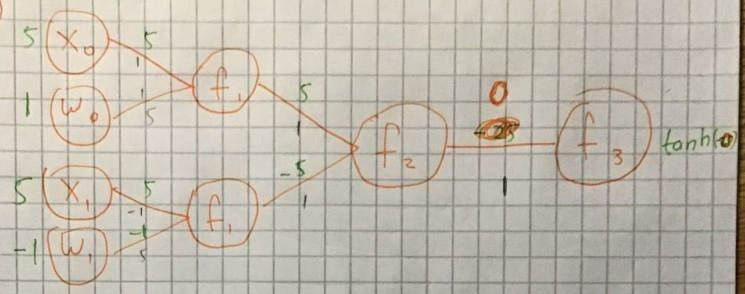
c)



	spam	not spam
buy	TP	TN
click	FP	FN

Q3:

a)



b) $\frac{\partial f_1}{\partial u} = c$ $\quad \frac{\partial f_3}{\partial u} = 1 - f_3(u)^2$

$\frac{\partial f_2}{\partial u} = 1$

c) ~~[illegible]~~ ans = $\tanh(0) = 0$

1 · 1 · -1

d) $\frac{\partial h}{\partial x_0} = \frac{\partial f_3}{\partial f_2} \cdot \frac{\partial f_2}{\partial f_1} \cdot \frac{\partial f_1}{\partial x_0} = 1$

$\frac{\partial h}{\partial w_0} = \frac{\partial f_3}{\partial f_2} \cdot \frac{\partial f_2}{\partial f_1} \cdot \frac{\partial f_1}{\partial w_0} = 5$

$\frac{\partial h}{\partial x_1} = \frac{\partial f_3}{\partial f_2} \cdot \frac{\partial f_2}{\partial f_1} \cdot \frac{\partial f_1}{\partial x_1} = -1$ $\quad \frac{\partial h}{\partial w_1} = \frac{\partial f_3}{\partial f_2} \cdot \frac{\partial f_2}{\partial f_1} \cdot \frac{\partial f_1}{\partial w_1} = 5$

Q4:

a) I would use a many-to-one recurrent neural network to feed through the video clips as multiple images in consecutive order, and this network will be able to read patterns in those images to produce a categorical result to describe the gesture.

b) use cost function $J(\theta)$: initialize θ, update over $\theta: \theta - \alpha \frac{\partial}{\partial \theta} J(\theta)$, until convergence.

c) Compared to other machine learning algorithms, deep learning can do feature engineering concurrently with its process because the deep learning layers are partly responsible for the feature engineering.

d) If you replace it with linear functions it will be harder to approximate non-linear functions because the non-linearity in the data cannot be easily described, and a non-linear function would do a better job at approximation.

e) Increase step size

Dropout

(Q5)

a) i) True Positive

ii) False Negative or False Positive

b) i) A mini-batch is a smaller chunk of the training data set used to compute the regularization of the gradient descent

ii) A mini batch is used because it helps increase efficency so that we don't have to compute all of the data in the dataset.

iii) [redacted]

back propagation $W \leftarrow W - \alpha \frac{\partial J(W)}{\partial W}$