Textbook questions for Deep Learning

1.

Overfitting

* Training sets are better than validation. Loss function not enough

2.

Logit (logistic)

* Measure of the probability that a guess is correct.
* Range -infinity to +infinity. If -infinity close to 0, if +infinity close to 1.
* Likelihood that the class belongs to class 1, given the input if the logit is 0.95

Loss function to measure logistic regression

* Cross entropy, binary cross entropy, multinomial

For other distributions

* Gaussian distrubtion = probit, general linear models.
* For probit regression use cross entropy still. Cross entropy is agnostic of distribution itself

In logistic regression, the frontier between the two classes

* In space – hyperplane

3.

RBF Kernel

* Most prevalent
* Gamma describes the affinity between two points in a space, if close the value is 1. If far away it is 0.
* Almost like close neighbours when the gamma is close to 1.

K-nearest neighbours

* Jagged lines correspond to the median lines between the different neighbours, tessellation of the space.
* If you have different shades you have different probabilities you have different neighbours.

Decision trees

* Vertical structures
* ADA boost means same shapes

Logistic Regression (SVM) with the kernel trick

* Distinguishing between polynomial and RDF
* Polynomial has iso lines like cones. Ellipses, hyperboles.
* Round blurbs are more like RDFs

4.

How many datasets do you need to have for training?

* 3, training, validation, test
* Training for training the algorithm.
* Validation to tune the parameters for the algorithm
* Test to assess the performance

Accuracy of 90% doesn’t mean anything.

* You need the confusion matrix. True positive, true false.
* In a medical sense, a test for a form of cancer. Small fraction of people will have it, need the split to show when a test is actually good or not.

5.

Feed forward neural net.

Neuron

Linear to reduce dimensionality. From n to 1. Linear combination makes sense.

Why do we use activation functions.

Normalisation layer – Batch Norm

Batch Norm

Why do we use normalisation layers.

What is vanishing gradients?