**Lecture 19 Poll**

**Slide 44**

**The portion of a network until the second-to-last layer (i.e. the layer just before the output layer) is essentially a “feature extraction” module that extracts linearly separable features for the classes, true or false?**

* True
* False

**The output layer is a linear classifier that can only perform well if the rest of network transforms the input space such that the classes are linearly separable, true or false**

* True
* False

**Slide 49**

**Select all that are true**

* A (classification) neural network is just a statistical model that computes the a posteriori probabilities of the classes given the inputs
* Training the network to minimize the KL divergence (Xentropy loss) is the same as maximum likelihood training of the network
* Training the network by minimizing KL divergence gives us a maximum likelihood estimate of the network parameters only when the classes are separable
* It is valid, and possibly beneficial, to train the network, and subsequently replace the final (output) layer by any other linear classifier

**Slide 72**

**An autoencoder with a linear activation in the hidden layer performs Principal Component Analysis of the input, True or False**

* True
* False

**An autoencoder with linear activations in the hidden layer, that has been trained on some data can only output values on the principal subspace of that data, regardless of the input**

* True
* False

**Slide 82**

**Select all that are true of autoencoders with non-linear activations**

* An autoencoder with nonlinear activation performs non-linear principal component analysis of the training data
* It finds the principal manifold (surface, which may not be linear) near which the training data lies
* The decoder of the non-linear AE can only generate data on the principal manifold of the training data regardless of the input
* The decoder of the non-linear AE is a “dictionary” which composes data like the training data, in response to any input