**VAE 1, lecture polls**

**Slide 16**

**Maximum-likelihood estimation of probability distributions is based on the theory that the world is a terribly boring place**

* True
* False

**Maximum-likelihood estimation estimates the values of the parameters of a probability distribution such that they maximize the probability of the training data**

* True
* False

**Slide 66**

**Mark all that are correct about the EM algorithm**

* It is an iterative algorithm that can be used to estimate probability distributions when the data are incomplete and have missing components or variables
* It iteratively maximizes an “ELBO” function with respect to model parameters
* It provides a closed form formula to estimate the parameters of the distribution

**Mark all that are true of the ELBO (Empirical Lower Bound) function**

* It is a lower bound on the actual log probability of the training data as computed by the model
* It is a function of the model parameters
* There are some settings of the model parameters where the ELBO can be greater than the log probability of the training data

**Slide 100**

**Select all that are true of EM estimation**

* In each iteration we “complete” the data, by filling in the missing components/variables, and estimate parameters from the entire completed data
* A data instance can be completed by filling in the missing terms with every possible value, in proportion to their a-posteriori probability, given the observed components of the data
* A data instance can be completed by randomly drawing samples of the missing components from their a-posteriori probability distribution, given the observed components of the data
* “Data completion” must be performed only once during the entire training (with EM)

**Slide 143**

**Select all that are true about PCA**

* PCA finds the principal subspace, such that approximating all training data by their projections onto this subspace results in the lowest error
* An optimal autoencoder with linear activations reconstructs all data as their projections on the principal subspace
* The bases of this subspace can be uniquely estimated without constraints
* One way to uniquely estimate the subspace is to require the bases of the subspace (the decoder weights of the AE) to be orthonormal
* Another way to estimate the subspace uniquely is to require the distribution of the latent variable Z to be standard Gaussian
* The decoder weights estimated using both above solutions will be the same