Self Assessment for Lecture Admission: Deep Learning

1. What properties does a Bayes optimal classifier imply?
(a) It classifies every pattern correctly.
(b) It minimizes the false positives.
(c) It achieves the maximum sample efficiency.
(d) It achieves the minimum possible error rate.

2. Regression generally refers to,
(a) estimating a linear model for the data.
(b) estimating a continuous function describing the data.
(c) estimating posterior probabilities from data.
(d) estimating class labels from data.

3. Features in Pattern Recognition should
(a) preserve class information while reducing dimensionality.
(b) increase class information while increasing dimensionality.
(c) preserve class information while increasing dimensionality.
(d) increase class information while reducing dimensionality.

4. Maximum Likelihood Estimation is a method,
(a) to choose a parametric model.
(b) to incorporate prior knowledge in parametric models.
(c) to classify data using a parametric model.
(d) to estimate parameters of a parametric model.

5. Under which assumptions is the nearest neighbour classifier using the L2 norm Bayes optimal?
(a) Bayes distributed with covariance matrices of unit norm and equal prior.
(b) Bayes distributed with identity covariance matrices and equal prior.
(c) Gaussian distributed data with covariance matrices of unit norm and equal prior.
(d) Gaussian distributed data with diagonal covariance matrices and equal prior.

6. Ensemble methods,
(a) select a strong classifier from sets of classfiers.
(b) train a weak classifier to become a strong classifier.
(c) use multiple weak classifiers to build a strong classifier.
(d) select features to build a strong classifier.

7. Supervised training means,
(a) a loss function is used.
(b) the loss function uses ground truth labels.
(c) the loss function is continuously monitored during training.
(d) a classifier is used to compute the loss.

8. Training sets are usually assumed to be IID. What does IID stand for?
(a) Independent and Identically Distributed.

(b) Indistinguishable and Ideally Distributed.
(c) Independent and Identically Distributed.
(d) Indistinguishable and Ideally Distributed.

9. Principal Components Analysis,
(a) preserves distances of a manifold.
(b) decorrelates features.
(c) performs a non-linear dimensionality reduction.
(d) prevents overfitting from happening.

10. Hidden Markov Models,
(a) are well suited to time series problems.
(b) are well suited if the dataset is imbalanced.
(c) are well suited to solving Markov Decision Processes.
(d) Are well suited to solving Reinforcement Learning Problems.

11. Minimizing a least squares problem using gradient descend
(a) Will find the same solution as any other method.
(b) Can lead to multiple local optima.
(c) Will fail to converge.
(d) Is usually faster than Pseudoinverse algorithms.

12. What is the Expectation Maximization algorithm used for?
(a) Maximum A Posteriori Estimation for models with sparse variables.
(b) Maximum A Posteriori Estimation for models with latent variables.
(c) Maximum Likelihood Estimation for models with latent variables.
(d) Maximum Likelihood Estimation for models with sparse variables.

13. How can we incorporate prior knowledge in the Expectation Maximization algorithm?
(a) Change the Expectation step.
(b) Change the Maximization step.
(c) Change the Expectation and Maximization steps.
(d) It is impossible.

14. Overfitting means,
(a) training does not converge.
(b) training converges very slow.
(c) we need to train on test-data.
(d) a method performs worse on test-data than on training-data.

15. Overfitting can be dealt with by:
(a) training on test-data.
(b) Applying Singular Value Decomposition on the results.
(c) Maximum A Posteriori estimation.
(d) Using more flexible classifiers.

16. What is a typical method to train a Gaussian Mixture Model?
(a) The Expectation Maximization algorithm.
(b) The Singular Value Decomposition.
(c) The Gradient Descend algorithm.
(d) Maximum Likelihood estimation.

17. What is Manifold learning?
(a) Methods fitting manifolds to data.

(b) Methods reducing dimensionality while assuming data lies on a low dimensional manifold.
(c) Methods to optimize parametric models using manifolds.
(d) Method to estimate performance of classifiers.

18. How can we practically obtain sparse solutions to optimization problems?
(a) Soft thresholding of the weights.
(b) L0-Norm constraint on the weights.
(c) L1-Norm constraint on the weights.
(d) L2-Norm constraint on the weights.

19. What is naïve about "Naïve Bayes"?
(a) Covariance matrices are assumed to be diagonal.
(b) Priors are assumed to be identical.
(c) Data is assumed to be Bayes distributed.
(d) No training is performed.

20. How is k-Nearest Neighbours different from k-Means?
(a) k-Nearest Neighbours can be optimized using Singular Value Decomposition.
(b) k-Means converges faster.
(c) k-NN is supervised while k-Means is unsupervised.
(d) k-NN uses the nearest samples for training while k-Means uses the means.