

Exercise sheet week 11 - Machine Learning I - 2015/16

Please send your submissions (runnable code, plots and written answers) **via email to weis@ccc.cs.uni-frankfurt.de until Friday January 22nd 2016**. One submission per student. Prepare to present your solutions in the exercise session. Students that are not able to explain their solutions may not be given credit on their submissions.

1 Regression in General (4 Points)

1. (Textual answer please) Describe in no more than five sentences what regression in general means.
2. (Textual answer please) What is the connection between regression and inference ?
3. (Textual answer please) In what cases would you use logistic regression instead of linear regression ?
4. (Textual answer please) What is the goal of the dual representation in logistic regression ?
5. (Textual answer please) What are the risks in applying regression on a small (probably noisy) dataset, and how could you prevent it ?
6. (Textual answer please) What is the connection between Occams Razor, the curse of dimensionality, and sparse linear regression ?
7. (Textual answer please) What are the benefits of using a Bayesian Model for regression ?
8. (Textual answer please) What is the *kernel trick* ?

2 Discriminative model - Classify faces (3 Points)

1. download `pca_faces.py` from the files section, run it
2. compute the score on training data
3. (Text answer please) Explain the process of whitening, why has it been applied in this case ?
4. (Text answer please) Describe the functionality of a support vector machine for classification in at most 5 sentences.
5. (Text answer please) Why do we first apply a PCA, when we use a Support Vector Machine after that for classification?

3 Generative model - Gaussian Mixture Models, supervised (3 Points)

1. Use `sklearn.datasets` to load the iris-dataset ¹
2. Use `sklearn.mixture` and create one GMM instance for every type of covariance matrix (spherical, tied, diag, full)
3. Use `sklearn.cross_validation` to apply a StratifiedKFold and separate your data in train/test sets.
4. (Text answer please) Why is this (StratifiedKFold) a great method to split your data ?
5. Use the `fit` function of your GMMs to estimate model parameters
6. Use the `predict` function of your GMMs to predict labels for the test-set
7. Calculate the accuracy of every GMM
8. (Text answer please) How and why do different types of covariance matrices influence the results ?

¹dataset = datasets.load_iris()

4 Generative model - Markov Chain text generation (2 Extra Points)

1. Download hmm.py and text.txt from the website
2. Explain in no more than five sentences what this program does
3. How are probabilities represented ?
4. Is this really a Markov chain ?
5. What can the model tell us about the data ?