Assignment 3

Machine Learning, Summer term 2015
Tobias Lang and Ulrike von Luxburg

To be discussed in exercise groups on April 27/29

Total number of points: 20

Exercise 1 (Linear mapping, 4 points (1+1+1+1))
Load the data from `Adot.mat`. Each column of matrix X represents one data point.

(a) Use the following code to calculate a linear mapping V. Apply the linear mapping on X to get $Y = VX$. Plot both X and Y in the same figure. What does the linear mapping V do?

```
theta = pi/3;
V = [cos(theta) -sin(theta);sin(theta) cos(theta)];
```

(b) Now apply the transpose of the linear mapping on Y to get $Z = V^t Y$. Plot Z and describe what the linear mapping $V^t V$ does.

(c) What do the linear mappings `D1=[2 0;0 2]` and `D2=[2 0;0 1]` do? Apply them on X and plot the results.

(d) What does the linear mapping $A = V^t * D2 * V$ do? Apply it on X and plot the result.

Exercise 2 (Estimating probabilities, 5 points (1+1+1+1+1))
In Assignment 1, you had to estimate probabilities from data. Probabilities of binary random variables can be modeled as the parameters p of a binomial distribution $B(n; p; k)$ where n is the number of trials, k is the number of successes and p is the success probability. For observed n and k, p can be estimated as $\hat{p} = \frac{k}{n}$. In this exercise, you are asked to empirically investigate the accuracy of such estimates.

(a) Take a look at the Matlab function `binornd`. For $n = 10$, sample success numbers for $p_1 = 0.4$ and $p_2 = 0.5$:

```
s1 = binornd(n, p1);
...
```

Estimate $\hat{p}_1$ and $\hat{p}_2$. What is the absolute and relative difference between the estimated and true probabilities? Is the order of the estimated probabilities the same as the order of the true probabilities ($\hat{p}_2 > \hat{p}_1$ and $p_2 > p_1$)?

(b) Run the previous experiment 100 times. Calculate the mean, maximum and minimum of the absolute and relative difference between the estimated and true probabilities. Calculate the percentage of runs in which the estimated probabilities are in the wrong order ($\hat{p}_2 < \hat{p}_1$).

(c) Run the experiment again 100 times, but now for $n \in \{100, 1000, 10000\}$. Visualize your results in different subplots (n on the x-axis in log-scale; mean/maximum/minimum absolute/relative differences and wrong-order-fractions on the y-axis).

(d) Repeat the whole experiment with (i) $p_1 = 0.1$ and $p_2 = 0.5$ and (ii) $p_1 = 0.01$ and $p = 0.02$.

(e) What do you conclude from your results about the accuracies of probability estimates?

Exercise 3 (Least-squares regression, 7 points (1+2+1+2+1)) In this exercise, you will implement the linear least-squares method for regression.

(a) Data preparation

- Load `reg1d.mat`. Plot training and test data.
- Preprocess the training data by concatenating 1 (for the bias term) to each training point.

(b) Learning

- Write a function `least_squares(X, Y)` which computes the weight vector w of the least-squares solution for input points $X \in \mathbb{R}^{n\times d}$ with target values $Y \in \mathbb{R}^{n\times 1}$. n denotes the number of points and d is the number of features (dimensions per point).
- Calculate w using `least_squares(X, Y)` for the given training data. Plot the prediction of the resulting classifier into your previous plot.

(c) Evaluation

- Write a function `err = lossL2(Y,Y_pred)` which returns the empirical squared error of predicting `Y_pred` instead of `Y`.
- What is the average L2 loss of the classifier on the test data?

(d) Non-linear features

- Add quadratic and cubic basis functions to your input features (add new colums for x^2 and x^3 in addition to 1 and x).
- Re-run learning and evaluation.

(e) Outlier

- Add an extreme outlier to your training data:

```
X_train = [X_train; 1.05];
Y_train = [Y_train; -10];
```

- Run your code to see its effect on linear least-squares regression.

Exercise 4 (Decision boundary, 4 points (1+3+1)) Generate a dataset by calling
`[X Y] = mixGaussian2d(100,0.4,0.6);`
X is a set of 2d points belonging to two different classes with labels in Y. As prior knowledge, you know that the density of each class is a Gaussian.

(a) Compute the sample mean and sample variance for each class.

(b) Using the estimated mean and variance from part (a), find the analytic decision boundary for the Bayes classifier. For help see Section 2.6.3 from the pattern recognition book by Duda et al. (available on the course web page).

(c) Plot this boundary and the sample points in a 2D plot.