Assignment 3

Machine Learning, Summer term 2014, Ulrike von Luxburg

To be discussed in exercise groups on May 5-7

Exercise 1 (Linear mapping, 1+1+1+1 points) 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 $Z = V^tY$. Plot Z and describe what does the linear mapping V^tV do.

(c) What does 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 (Inverse of a matrix, 1 point) Assume that V is a $n \times n$ matrix such that $VV^t = V^tV = I$, where I is the identity matrix. Moreover, D is a diagonal matrix

$$D = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix} \tag{1}$$

where $d_i > 0$. Prove that the inverse of the matrix $A = VDV^t$ is $A^{-1} = VD^{-1}V^t$. Here, D^{-1} is a diagonal matrix with diagonal entries $1/d_i$. To prove that $B = A^{-1}$, it is enough to show that $AB = BA = I$.

Exercise 3 (Convexity, 2 points) Prove that the least squares loss function $\|Y - Xw\|^2$ is a convex function of w.

Exercise 4 (Linear regression with weights, 4 point) Consider a data set in which each data point X_i is associated with a weighting factor $r_i > 0$, so that the empirical least squares error becomes

$$E = \frac{1}{n}\sum_{i=1}^{n} r_i(Y_i - \langle w, X_i\rangle)^2.$$

Find an expression for the solution w that minimizes this error function.

Exercise 5 (Ridge regression, 2+2+4 points) In this exercise, you will implement the ridge regression algorithm. Load the synthetic train and test data from `dataRidge.mat`.

(a) Run the linear least square and plot the training points, the predicted line and the predicted values for the test data.

(b) Linear least square with polynomial basis functions

Web page: `http://www.informatik.uni-hamburg.de/ML/contents/people/luxburg/teaching/2014-ss-vorlesung-ml/`
Login: “machine”, Password: “learning”

- Run the linear least square with polynomial basis functions

$$\Phi_i(x) = x^i; i = 1, ..., 15. \tag{2}$$

  Illustrate the learned regression function by applying it on `xx=-1.5:0.01:2.5`.
- Describe the class of functions that you can learn with these basis functions.

(c) Ridge regression

- Write a function `RidgeLLS(X,Y,lambda)` which implements the ridge regression. Here, X is the design matrix.
- Apply the ridge regression on the test data using the set of basis functions in Equation 2. Plot the prediction function (on the previous figure) for regularization constant $\lambda \in \{0.0001, 0.1, 10\}$.
- Report the prediction error with respect to λ for $\lambda \in \{2^i; i = -15, -14, ..., 1\}$.

Exercise 6 (Prediction complexity of linear least square and kNN, 1 point) Compare the prediction running time for linear least square method and kNN regression (computational complexity). How much information do you need to keep for predicting with each method (space complexity)?