Assignment 4

Machine Learning, Summer term 2015

Tobias Lang and Ulrike von Luxburg

To be discussed in exercise groups on May 4/6

Total number of points: 20

Exercise 1 (Ridge regression, 1+1+5 points) In this exercise, you will implement the ridge regression algorithm. Load the synthetic training and test data from `dataRidge.mat`.

(a) Calculate the linear least-squares solution for the training data. Plot the training and test points and the predicted line.

(b) Calculate the linear least-squares solution using the polynomial basis functions

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

Illustrate the learned regression function by applying it on `xx=-1.5:0.01:2.5`.

(c) Ridge regression

- Write a function `RidgeLLS(X,Y,lambda)` which implements the ridge regression. Here, X is the design matrix.
- Calculate the rigde regression solution from the training data using the set of basis functions in Equation 1 and different regularization constants $\lambda \in \{0, 0.0001, 1, 1000\}$.
- Predict the test data. Plot the prediction function (in the previous figure) by applying it on `xx=-1.5:0.01:2.5`.
- Report the prediction error (L2-loss) with respect to λ for $\lambda \in \{2^i; i = -15, -14, ..., 1\}$ and $\lambda = 0$.

Exercise 2 (Prediction complexity of linear least square and kNN, 1 point) Compare the prediction running time for linear least-squares and kNN regression (computational complexity). How much information do you need to store for predicting with each of these methods (space complexity)?

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

Exercise 4 (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

$$D = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix} \tag{2}$$

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 $AB = BA = I$.

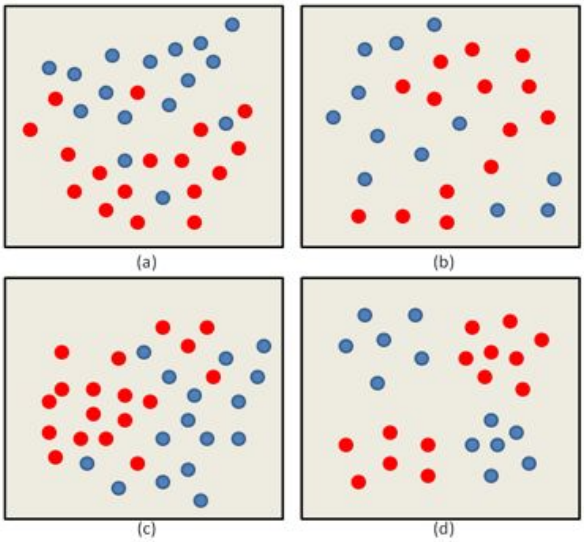
Exercise 5 (Least-squares regression for multi-class classification, 6 points (2+2+2)) The Iris flower data-set consists of 50 samples from each of three species (Iris setosa, Iris virginica and Iris versicolor). Four features were measured from each sample: the length and the width of the sepals and petals, in centimeters. Your task is to predict the species from these features.
A simple approach to multi-class classification is to divide the problem into binary subproblems. Each subproblem deals with the classification for one of the classes c (class c vs. not c). The corresponding target vector is $y_c \in \{0,1\}^n$ (for n examples) where $y_{c,i} = 1$ if x_i belongs to c and $y_{c,i} = 0$ otherwise. In this exercise, we approximate each binary classification problem with regression (in other words, we regress on only two values, namely 0 and 1). For each class c, we calculate the least-squares regression function f_c from X and y_c. Based on all $\{f_c\}$, we predict $y_{pred} = \text{argmax}_c f_c(x)$; this is the class of the regression function with the largest value. For example, for $f_1(x) = 0.31$, $f_2(x) = 0.39$ and $f_3(x) = 0.17$, we predict label 2. (Alternatively, we could fit a single least-squares model for all classes simultaneously, see Sec. 4.2 in the book of Hastie et al..)

1. Load `iris_multiclass.mat`. You have 3 classes in your dataset. Estimate w_{setosa}, $w_{versicolor}$ and $w_{virginca}$ using linear least-squares based on the training data. Note that you have to create appropriate training and test data yourself from `indices_train` and `indices_test`.

2. Predict the classes for the test data (helpful command: `max`). Report the misclassification error (0/1 loss).

3. What are potential problems of using least-squares regression for multi-class classification? Take a look at Sec. 4.2 in the book of Hastie et al..

Exercise 6 (Decision boundary, 1+2 point) Consider the following four samples, where colors indicate class labels. Each sample is typical of an underlying distribution.



Plot the decision boundaries for the four samples when solving this classification problem with:

(a) a 1NN classifier;

(b) linear least-squares regression without and with quadratic basis functions.