Machine Learning for Robotics SS 2017
Instructor: Georg Martius <georg.martius@tuebingen.mpg.de>
Due-date: 04.05.2017 (hand in at beginning of recitation session)
Exercise Sheet 2
April 27, 2017

1 Bayes Theorem

(a) 1. You go to see the doctor about an ingrowing toenail. The doctor selects you at random to have a blood test for swine flu, which for the purposes of this exercise we will say is currently suspected to affect 1 in 10,000 people in Germany. The test is 99% accurate, in the sense that the probability of a false positive is 1%. The probability of a false negative is zero. You test positive. What is the new probability that you have swine flu?

(b) Now imagine that you went to a friends wedding in Mexico recently, and (for the purposes of this exercise) it is know that 1 in 200 people who visited Mexico recently come back with swine flu. Given the same test result as above, what should your revised estimate be for the probability you have the disease?

(c) Imagine that, while in Mexico, you also took a side trip to Las Vegas, to pay homage to the TV show CSI. Late one night in a bar you meet a guy who claims to know that in the casino at the Tropicana there are two sorts of slot machines: one that pays out 10% of the time, and one that pays out 20% of the time [note these numbers may not be very realistic]. The two types of machines are coloured red and blue. The only problem is, the guy is so drunk he cant quite remember which colour corresponds to which kind of machine. Unfortunately, that night the guy becomes the vic in the next CSI episode, so you are unable to ask him again when hes sober. Next day you go to the Tropicana to find out more. You find a red and a blue machine side by side. You toss a coin to decide which machine to try first; based on this you then put the coin into the red machine. It doesnt pay out. How should you update your estimate of the probability that this is the machine youre interested in? What if it had paid out - what would be your new estimate then?

2 Model selection and Cross validation

Last week you implemented linear regression. Now we want to use this code to gain experience with model selection. We are again using the `diabetes.txt` dataset from the course website. Use the first 10 columns as features and the last column as target value.

(a) run the following experiment 50 times with different random seeds:

- do a random split of the dataset to choose 20 data points as training and the 20 points as validation set
- train a least squares regression model without regularization (using matrix operations)
- for each regularization parameters $\lambda \in \{2^{-20}, 2^{-19}, \ldots, 2^{10}\}$ train a ridge regression model with regularization λ (using matrix operations)
- evaluate the trained models on the validation set
- plot the results in a style that you find most appropriate/informative.

(b) What do you observe?

(c) Implement a K-fold cross validation.

- Take a set of 242 points aside for testing, which leaves you with 200 training points.
- Implement K-fold cross validation. Evaluate the performance of K=1,2,5 and 10 on the test-set.