

DM825 - Introduction to Machine Learning

Sheet 4, Spring 2013

Exercise 1

Repeat Exercise 3 of Sheet 3, replacing the multinomial distribution with an arbitrary exponential family distribution, and the Dirichlet distribution with the corresponding exponential family conjugate distribution. You are to show that in general the predictive probability $p(x_{new}|x_1, x_2, \dots, x_N)$ is a ratio of normalizers.

Exercise 2

Task 2 of Exam 2010.

Exercise 3

Examine the iris data example in R.

```
#  
?iris # read documentation  
data(iris) # load data  
str(data) # examine contents
```

As you see there are 5 features that can be used to predict one of the 3 Species. You can get some visualization of the data via `example(matplot)`.

If you do not like this famous example from botanic you can use the wine data set from the UCL repository or any similar data for classification task of your choice.

Fit a GLM model in R and assess its prediction error via cross validation. (Use `mlogit` from the homonymous package.)