IN4320 Machine Learning
Assignment Covariate Shift

Wouter M. Kouw

March 29, 2017

In order to complete this assignment, please provide a max. 2-page report that answers the following 3 questions and contains the 2 output figures described in the Code Assignment section.

1 Questions

Please answer the following questions related to the lecture on covariate shift.

1.1 (2 points)

Please find a real-life example of a domain adaptation or a covariate shift problem. Clearly identify what you believe corresponds to the source and target domain and motivate your argument on why the class-posterior distributions are equal.

1.2 (2 points)

Why is using a ratio of probability distributions as importance weights problematic?

1.3 (2 points)

Why is the sample average of the weights constrained to be close to 1 in the non-parametric estimator?

2 Code Assignment

On blackboard, you can find MATLAB-code that generates an artificial covariate shift setting and trains both a standard least-squares and a weighted least-squares classifier on the data. If you want, you can use a different programming language, but then you should translate the provided MATLAB -code to your language of choice in order to construct the deliverable.

Goal Your job is to implement Kernel Mean Matching (write in KMM.M) that builds the weighted classifier. Consult the lecture and the reading material on Blackboard for help.

Optimization You do not have to write a quadratic program solver yourself, you are welcome to use native libraries such as MATLAB's optimization toolbox (specifically, quadprog). If you want to make use of a different programming language, look for native implementations of a quadratic program solver (such as scipy's optimization package for Python), or use external tools such as CVXOPT (http://cvxopt.org/) or Gurobi (http://www.gurobi.com/).

Result Please provide two figures from which we can check whether you have successfully implemented KMM and the weighted classifier.

- For the first figure, please include a histogram of the values of the importance weights that you found (2 points).
- For the second figure, please provide a visualization of both the unweighted classifier and the weighted classifier on both the source and the target domain (2 points).