Supervised Learning 2004 — Assignment 5

Fernando Pérez-Cruz fernando@gatsby.ucl.ac.uk

Due Wednesday 15th December

Please hand in Q1, the comments in Q2 and the tree for Q3. Send by e-mail (subject Assignment 5) with the proposed code to solve Q3 and the figure in Q2.

Support Vector Machine for regression

Obtain the dual of the following SVM formulation for regression:

$$\min_{\mathbf{w},b,\xi_i,\xi_i^*} \left\{ \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \right\}$$

and

$$\min_{\mathbf{w},b,\xi_i,\xi_i^*} \left\{ \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\left((\xi_i)^2 + (\xi_i^*)^2\right) \right\}$$

both of them subject to the standard constraints:

$$\begin{aligned} y_i - (\boldsymbol{\phi}^\top(\mathbf{x}_i)\mathbf{w} + b) &\leq \epsilon + \xi_i \\ (\boldsymbol{\phi}^\top(\mathbf{x}_i)\mathbf{w} + b) - y_i &\leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* &\geq 0 \end{aligned}$$

Both dual must depend only on the Lagrange multipliers for the first 2 constraints (α_i and α_i^*). Must be expressed in terms of the kernel of $\boldsymbol{\phi}(\cdot)$.

Also show how b can be computed from the solution in terms of α_i and α_i^*.

Q2 Kernel PCA

Obtain the SVM solution with an RBF kernel for the provided data. (you can use the `irwls_svc.m` code provided in the demo package on the 29^{th} of November).

Report the solution in a 2D plot using `svcplot.m` and the used value of σ and C for the SVM.

Use the `kpca.m` function to extract the principal components of the data using the same kernel parameters as before. Solve the classification problem using an increasing number of components from the KPCA expansion.

a) How many components will you need to get the same solution as in the previous solution?

b) How many components will give you an approximate solution that does not differ significantly?

c) Which are the advantages and disadvantages of each method?

Add any comments that you believe relevant to explain you results.

Q3 Decision Trees

Using the connect-4 data set, compute the features for the decision tree showed in the figure. Also report the probabilities of winning, drawing and losing in each one of the leaves in the tree. Compare them to the prior probabilities of winning, drawing and losing, do you thing the decision tree give you some information about who is going to win after seeing 3 features? Explain you answer.