

COMP4702/COMP7703 - Machine Learning

Homework 8 - Kernel Machines

Marcus Gallagher

Core Questions

1. (From Marsland's text) Suppose that the following are a set of points in two classes:

$$\text{class 1 : } \begin{pmatrix} 1 \\ 1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

$$\text{class 2 : } \begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Plot them and find the optimal separating line. What are the support vectors, and what is the margin¹? Note: this question can be done “by hand”/using “pen and paper”.

2. Using the data and your answer from the previous question, calculate the (Euclidean) distance between the optimal separating line (i.e. the discriminant) and the point $(5.23 \ 7.14)^T$.
3. Lodhi et al. [1] describe the string subsequence kernel and illustrate its calculation with an example (p.422-423). If this example was expanded to include the word “rat”, how many dimensions would the feature space have?

Extension Questions

4. From the lecture material, we know that there is an equivalence between the kernel function and the basis functions in support vector machines:

$$k(\mathbf{x}, \mathbf{x}') = \phi(\mathbf{x}) \cdot \phi(\mathbf{x}')$$

for two data points \mathbf{x} and \mathbf{x}' . One interesting example of a kernel function is the Arc-cosine kernel:

$$k_{ac}(\mathbf{x}, \mathbf{x}') = \frac{\|\mathbf{x}\| \|\mathbf{x}'\|}{2\pi} (\sin \theta + (\pi - \theta) \cos \theta)$$

which is known to have an equivalence

$$k_{approx}(\mathbf{x}, \mathbf{x}') = \frac{1}{n} \sum_{i=1}^n \max(0, \mathbf{w}_i \cdot \mathbf{x}) \max(0, \mathbf{w}_i \cdot \mathbf{x}') \approx k_{ac},$$

¹That is, the distance from the discriminant to the nearest data point on one side. Some books may define the margin as twice this distance (i.e. both sides).

with the approximation becoming exact as $n \rightarrow \infty$. Here, \mathbf{w}_i is a weight vector of dimensionality equal to the dimensionality of \mathbf{x} , with each entry in \mathbf{w}_i drawn independently from $\mathcal{N}(0, 1)$.

Examine this approximation by writing a program to compute and plot the value of k_{ac} and k_{approx} as a function of θ .

Hints:

- Use the dataset supplied (hw8.csv) as your input data. Use (e.g) the first 500 columns of the dataset as examples of \mathbf{x} , the other half as examples of \mathbf{x}' . The angles between these vectors (θ) have been calculated for you in the final column. Note also that in this dataset, $\|\mathbf{x}\| = \|\mathbf{x}'\| = 1$ for all data points.
 - Plot k_{ac} as a curve, and overlay k_{approx} as points that should roughly follow the curve.
 - State the value of n you have used to create your plot.
5. Consider k_{approx} from the previous question. What machine learning model (recently covered in the course!) does this equation describe? Be specific about the model (e.g. its structure and any functions used within the model).

References

- [1] Lodhi, H., Saunders, C., Shawe-Taylor, J., Cristianini, N., Watkins, C. (2002). Text classification using string kernels. *Journal of Machine Learning Research*, 2(Feb), 419-444.