MATH 637 – Homework 5

Due: Thursday, May 9th, 5:15pm

This homework has two components: a written part and a simulation part

- Submit your solutions to the written part to Canvas as a PDF file. You may scan (or take a good picture of) a handwritten document, but they will be returned ungraded if they are not legible.
- Submit your solutions to the simulation part in a Colab notebook (which contains the corresponding codes and figures).

1 Written part (3%)

Suppose that the data set $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$ are generated by the following rules:

- $x_i = (x_i^{(1)}), x_i^{(2)})$ are chosen uniformly at random on the domain $[-3, 3] \times [-3, 3]$
- the label y_i is computed by

$$y_i = \text{sign}\left((x_i^{(2)})^2 - (x_i^{(1)})^3 + 2(x_i^{(1)}) - 1\right)$$

Construct a feature map $\phi : \mathbb{R}^2 \to \mathbb{R}^p$ for some $p > 2$ such that the mapped dataset $\{(x_1', y_1), (x_2', y_2), \ldots, (x_n', y_n)\}$ where

$$x_i' = \phi(x_i)$$

is linearly separable (i.e., separable by a hyperplane in $\mathbb{R}^p$). Write down explicitly the equation of the separating hyperplane for this classification dataset.

2 Simulation part

The goal of this part is to implement the feature map constructed in the written part explicitly.

(1%) Step 1. Generate dataset

Write Python code to generate a training dataset (x, y) and a test dataset (x_{test}, y_{test}) (following the rules described in the previous part) that contain 10,000 examples and 5,000 examples, respectively.

(1%) Step 2. Implement the feature map ϕ from the written part as a Python function.

(3%) Step 3 (Predictions from the mapped feature space)

- Use the function in Step 2 to obtain the transformed training dataset $\{(x'_1, y_1), (x'_2, y_2), \ldots, (x'_n, y_n)\}$ where
$$x'_i = \phi(x_i)$$
- Construct a linear SVM classification model F on the transformed dataset (x', y)
- Predict the outputs of your model $F(\phi(x))$ on the test set. Compute and print out the accuracy of this model.

(2%) Step 4 (Predictions from the original space using kernel SVM)

- Construct a kernel SVM classification model G on the original training dataset (x, y) using
  - polynomial kernel with degree d
  - regularization parameter C
  - $coef0 = 1.0$

  where d and C are chosen by cross-validation over a reasonable set of candidates.
- Predict the outputs of your model $G(x)$ on the test set. Compute and print out the accuracy of this model.