Submission Date: 15th November 2022

1 Support Vector Machines

1. Explain the "kernel trick" and why we use it in SVMs.
2. What is the difference between hard- and soft-margin SVM?
3. The SVM problem can be considered as optimizing the balance between the average hinge loss over the examples and the regularization term that keeps the parameters small (increasing the margin). This balance is set by the regularization term $\lambda > 0$. Here, we focus on the case without the offset parameter w_0 (setting it to zero), hence the training objective will become:

$$\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}(y_i, w^T x_i) + \frac{\lambda}{2}\sum_{j=1}^{m} w_j^2$$

   where the hinge loss is given by

$$\mathcal{L}(y_i, w^T x_i) = \max(0, 1 - y_i(w^T x_i))$$

   (a) Compute the gradient of the training objective w.r.t. w and describe the pseudocode for the gradient descent algorithm.

   (b) Use gradient descent to update the weights based on the following example:

$$\lambda = 0.5, \eta = 0.01, y = 1, x = \begin{bmatrix} 1.0 \\ 0.0 \end{bmatrix}, w = \begin{bmatrix} 1.0 \\ 1.0 \end{bmatrix}$$

2 Linear Separability

Given the following dataset:

x	y
-3	-1
-2	-1
-1	-1
0	1
1	1
2	-1
3	-1

1. Create a sketch of the data. Is it linearly separable? If so, draw a separating hyperplane.

2. Apply the mapping $g : \mathbb{R} \to \mathbb{R}^2$ defined by:

$$g(x) = \begin{pmatrix} x \\ x^2 \end{pmatrix} \tag{1}$$

on all the data points to create the transformed data set and create a plot of it. Is the transformed data set linearly separable? If yes, find a separating hyperplane $< \mathbf{w}, \mathbf{x} > +b = 0$, compute its parameters[1] and plot it.

3. Use the computed hyperplane from 2. and compute its output for $x = \frac{1+\sqrt{5}}{2}$. Does it belong to the positive or negative class and why?

[1] It is fine if you use the border points only, i.e. only active constraints