AU 311 Introduction to Pattern Recognition

By: Harry Lording (717030990012)

HW#: 2

December 13, 2019

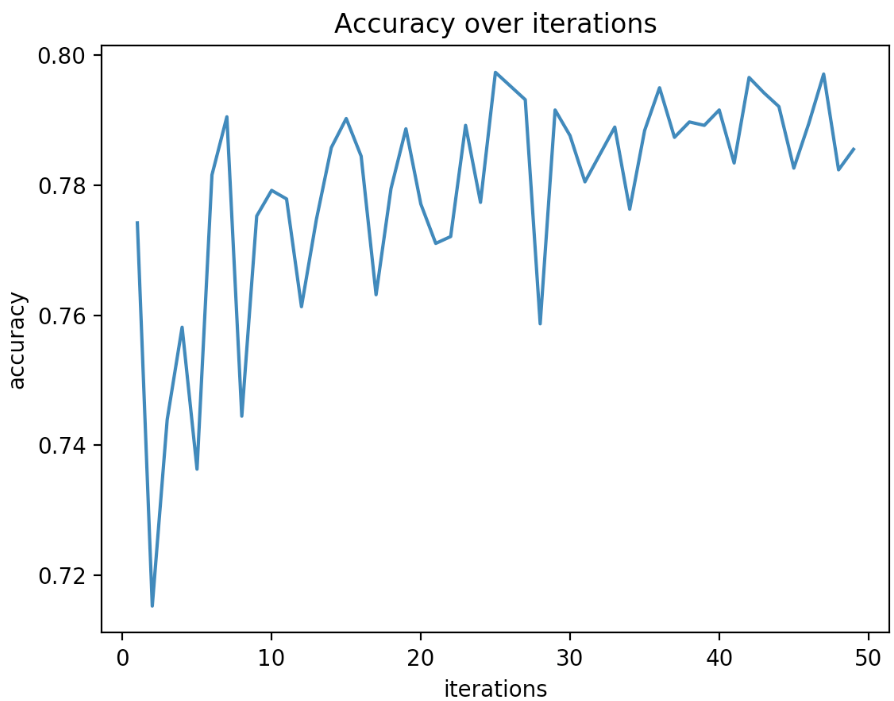
Problem 1

Solve the following linear l2-SVM by SGD method,

$$
\begin{aligned}
\min_{w,\rho,\xi} \quad & \frac{1}{2}\|w\|_2^2 + \frac{1}{m}\sum_{i=1}^{m}\xi_i^2 \\
\text{s.t.} \quad & y_i(w^T x_i + b) \geq 1 - \xi_i \\
& \xi_i \geq 0, \forall i = 1, 2, \ldots m.
\end{aligned} \tag{1}
$$

Try your code on dataset "magic04" (data provided, the last column stands for the label).

i) report the classification accuracy on the test data and plot the training accuracy v.s. the SGD iteration.



Accuracy over the amount of iterations of SGD. After roughly the 25 iterations of the samples, the SVM converges to an accuracy just under 0.8.

ii) numerically find the best ratio of samples when calculating the SGD. (For example, to achieve certain accuracy with the shortest time.)

The graph tracks the split of split data with training data, essentially showing the ratio of training samples needed to converge. As expected when the test split consists of only 0.01 of the data the SVM will perform best (it seems like the SVM didn't overtrain on the data). However it seems at around a test split ratio of 0.6 test data to 0.4 training data the SVM reached almost its best accuracy of 0.78. Therefore the best ratio for the SVM to efficiently reach it's best performance is training on roughly 0.4 ratio of the samples.

Problem 2

There have been many variants of SVM for different purpose. The following is called ν-SVM which can controls the ratio of support vectors. The primal formulation of ν-SVM is given as

$$\begin{aligned} \min_{w,\rho,\xi} \quad & \frac{1}{2}\|w\|_2^2 - \nu\rho + \frac{1}{m}\sum_{i=1}^{m}\xi_i \\ \text{s.t.} \quad & y_i(w^T x_i + b) \geq \rho - \xi_i \\ & \rho \geq 0, \xi_i \geq 0, \forall i = 1, 2, \ldots m. \end{aligned} \tag{2}$$

Please derive its dual problem and discuss the meaning of ν.

Working out for the question completed in the photos below.

$$\min_{w,\rho,\xi} \quad \frac{1}{2}\|w\|_2^2 - v\rho + \frac{1}{m}\sum_{i=1}^{m} \xi_i^2$$

$$\text{st} \quad y_i(w^T x_i + b) \geq 1 - \xi_i$$

$$\xi_i \geq 0, \quad \forall i = 1, 2 \ldots m.$$

~~To simplify the dual, as L(~~

First derive the Lagrangian ~~formula~~ ~~equa~~ of the equation using multiples $\alpha_i, \beta_i, \delta \geq 0$.

$$L(w, \xi, b, \rho, \alpha, \beta, \delta) = \frac{1}{2}\|w\|^2 - v\rho + \frac{1}{m}\sum_{i=1}^{m} \xi_i$$

$$- \sum_{i=1}^{m} (\alpha_i(y_i(\langle x_i, w\rangle + b) - \rho + \xi_i) + \beta_i \xi_i - \delta\rho).$$

$$= \frac{1}{2}\|w\|^2 - v\rho + \frac{1}{m}\sum_i \xi_i - \delta\rho - \sum_i \xi_i \beta_i - \sum_i \alpha_i (y_i w x_i + \xi_i - \rho)$$

This function needs to be minimized according to the primal variables w, ξ, b, ρ and maximised by the dual variables α, β, δ.

To simplify $L(w, \xi, b, \rho, \alpha, \beta, \delta)$, as the function is convex when α, β, δ fixed, for any given α, β, δ

$$\frac{\partial}{\partial b} L(w, \xi, b, \rho, \alpha, \beta, \delta) = 0,$$

$$\frac{\partial}{\partial w} L(w, \xi, b, \rho, \alpha, \beta, \delta) = 0$$

$$\frac{\partial}{\partial \xi} L(w, \xi, b, \rho, \alpha, \beta, \delta) = 0$$

and $$\frac{\partial}{\partial \rho} L(w, \xi, b, \rho, \alpha, \beta, \delta) = 0.$$

Solving the following partial derivatives we obtain.

$$w = \sum_{i=1}^{m} \alpha_i y_i x_i$$

$$\alpha_i + \beta_i = 1/m$$

$$\sum_{i=1}^{m} \alpha_i y_i = 0$$

$$\sum_{i=1}^{m} \alpha_i - \delta = v.$$

substituting $w = \sum_{i=1}^{m} \alpha_i y_i x_i$ and $\frac{1}{m} = \alpha_i + \beta_i$ into $L(\alpha, w, \xi, \rho, \beta, \delta)$ ~~we get~~ and given the fact that $\frac{1}{m} - \alpha_i + \beta_i = 0$ we can reduce the function to

$$\max_{\alpha} \quad -\frac{1}{2}\sum_{ij} \alpha_i \alpha_j y_i y_j x_i' x_j$$

$$\text{s.t} \quad 0 \leq \alpha_i \leq \frac{1}{m}$$

$$\sum_i \alpha_i \geq v$$

) if the property $\rho > 0$ holds then it can be said that v is an upper bound on the equation $\frac{\text{number of margin errors}}{m}$

and v is a lower bound on:

$$\frac{\text{number of support vectors} + \text{number of margin}}{m}$$