HW 5

Arnab Dey
Student ID: 5563169
Email: dey00011@umn.edu

Solution 1.a

Fig. 1 shows the decision boundary for the dummy data created in Q1.a. The error rate on this dataset is 0%.

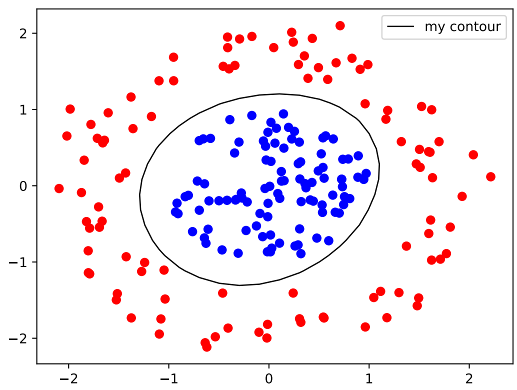

Figure 1: Classification with polynomial kernel of degree 3

Solution 1.b

Fig. 2 shows the decision boundary returned by 'svc' function of 'SVM' module of 'sklearn' package with the same kernel as previous one. Fig. 2 shows that the decision boundary returned by 'sklearn' is better

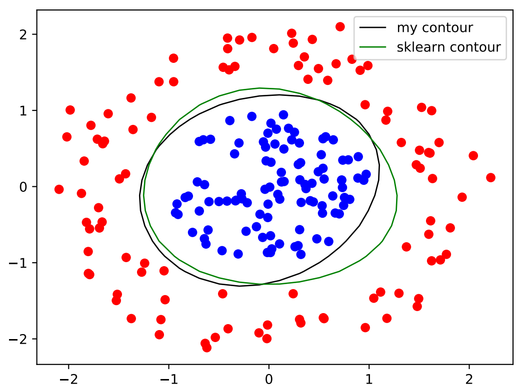

Figure 2: Classification with polynomial kernel of degree 3 with scikit-learn module

because of the additional regularization parameter used by 'sklearn' package. Fig. 2 shows the decision boundary for regularization parameter $C = 0.05$. Fig. 3 shows that decreasing the regularization parameter makes the decision boundary wider thus causing more error on training set. On the other hand, increasing the regularization parameter shrinks the decision boundary and too much increment will also cause larger training error.

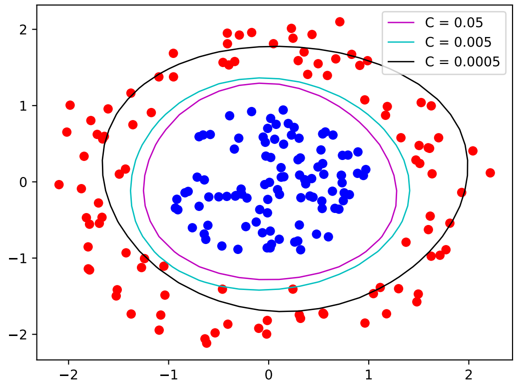

Figure 3: Classification with polynomial kernel of degree 3 with different regularization parameter C of scikit-learn module

Solution 1.c

The following table shows the training and test error rates on both 'optdigits49' and 'optdigits79' dataset.

Dataset	Training error (%)	Test error (%)
optdigits49	0.473	3.169
optdigits79	0.355	0.709

Table 1: Q1.c: Error-rate on subset of optdigit dataset

Solution 2

Hard margin version of the ν-SVM is defined by the following optimization problem:

$$\begin{aligned} &\underset{\boldsymbol{w}}{\text{minimize}}\ \frac{1}{2}\|\boldsymbol{w}\|^2 - \nu\rho \\ &\text{subject to } r^t(\boldsymbol{w}^T\boldsymbol{x}^t + w_0) \geq \rho \\ &\rho \geq 0. \end{aligned}$$

We first write the unconstrained problem using Lagrange multipliers α^t and μ^t as follows:

$$L_p = \frac{1}{2}\|\boldsymbol{w}\|^2 - \nu\rho - \sum_t \alpha^t[r^t(\boldsymbol{w}^T\boldsymbol{x}^t + w_0) - \rho] - \sum_t \mu^t\rho \tag{1}$$

(1) has to be minimized w.r.t $\boldsymbol{w}, w_0$. The dual problem is to maximize L_p w.r.t α^t subject to the constraint that the gradient of L_p w.r.t $\boldsymbol{w}, w_0$ and ρ is 0. Therefore,

$$\begin{aligned} \frac{\partial L_p}{\partial \boldsymbol{w}} &= \boldsymbol{w} - \sum_t \alpha^t r^t \boldsymbol{x}^t = 0 \implies \boldsymbol{w} = \sum_t \alpha^t r^t \boldsymbol{x}^t \\ \frac{\partial L_p}{\partial w_0} &= -\sum_t \alpha^t r^t = 0 \implies \sum_t \alpha^t r^t = 0 \\ \frac{\partial L_p}{\partial \rho} &= -\nu + \sum_t \alpha^t - \sum_t \mu^t = 0 \implies \sum_t \alpha^t = \nu + \sum_t \mu^t \end{aligned} \tag{2}$$

Since $\mu^t \geq 0$, the last implies $\sum_t \alpha^t \geq \nu$. Plugging the above into (1), we get the dual L_d,

$$\begin{aligned}
L_d &= \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} - \nu\rho - \sum_t [\alpha^t r^t \boldsymbol{w}^T \boldsymbol{x}^t + \alpha^t r^t w_0 - \alpha^t \rho] - \sum_t \mu^t \rho \\
&= \frac{1}{2}\sum_t \sum_s \alpha^t \alpha^s r^t r^s (\boldsymbol{x}^t)^T \boldsymbol{x}^s - \nu\rho - \sum_t \sum_s \alpha^t \alpha^s r^t r^s (\boldsymbol{x}^t)^T \boldsymbol{x}^s + \sum_t \alpha^t \rho - \sum_t \mu^t \rho \\
&= -\frac{1}{2}\sum_t \sum_s \alpha^t \alpha^s r^t r^s (\boldsymbol{x}^t)^T \boldsymbol{x}^s \\
&\text{subject to } \sum_t \alpha^t r^t = 0, \ \sum_t \alpha^t \geq \nu
\end{aligned} \tag{3}$$