# ECE: 523 HW #3

## David Akre

## Theory

<u>Support Vector Machines</u>

David Akre | ECE:523 HW#3 | 3/3/18 | 1

(1) <u>Support Vector Machine</u>: In class we discussed that if our data is not linearly separable, then we need to modify our optimization problem to include slack variables. The formulation that was used is known as the $l_1$-norm soft margin SVM. Now consider the formulation of the $l_2$-norm soft margin SVM, which squares the slack variables with the sum. Notice the non-negativity of the slack variables has been removed:

$$\arg\min \frac{1}{2}\|w\|_2^2 + \frac{c}{2}\sum_{i=1}^{n}\varepsilon_i^2 \quad \text{s.t. } y_i(w^Tx_i + b) \geq 1 - \varepsilon_i \quad \forall i \in [n]$$

Derive the dual form expression along with any constraints. Work must be shown.

Solution is given by:

$$\arg\max \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j x_i^T x_j - \frac{1}{2c}\sum_{i=1}^{n}\alpha_i^2 \quad \text{s.t. } \alpha_i \geq 0 \ \forall i \in [n] \text{ and } \sum_{i=1}^{n}\alpha_i y_i = 0$$

---

<u>Given</u>: $\arg\min \frac{1}{2}\|w\|_2^2 + \frac{c}{2}\sum_{i=1}^{n}\varepsilon_i^2$ s.t. $y_i(w^Tx_i + b) \geq 1 - \varepsilon_i \quad \forall i \in [n]$

- $l_2$-norm soft margin: $\frac{1}{2}\|w\|_2^2$
- Sum of square slack variables: $\frac{c}{2}\sum_{i=1}^{n}\varepsilon_i$ where $c$ = defines how large or small the margin (constant) will be
- Constraint: $y_i(w^Tx_i + b) \geq 1 - \varepsilon_i$

<u>Step 1</u>: Form a lagrangian s.t. constraints

$$L(\alpha, \mu, w) = \frac{1}{2}\|w\|_2^2 + \frac{c}{2}\sum_{i=1}^{n}\varepsilon_i^2 - \sum_{i=1}^{n}\mu_i\varepsilon_i - \sum_{i=1}^{n}\alpha_i[y_i(w^Tx_i + b) - 1 + \varepsilon_i]$$

<u>Step 2</u>: Take partial derivatives of $L(\alpha)$

- $\frac{dL}{d\varepsilon_i} = c\varepsilon_i - \mu_i - \alpha_i = 0 \rightarrow \overset{(1)}{c = \frac{\mu_i + \alpha_i}{\varepsilon_i}}$
- $\frac{dL}{db} = \overset{(2)}{\sum_{i=1}^{n}\alpha_i y_i = 0} \leftarrow$ New constraint
- $\frac{dL}{dw} = w - \sum_{i=1}^{n}\alpha_i y_i x_i = 0 \rightarrow \overset{(3)}{w = \sum_{i=1}^{n}\alpha_i y_i x_i}$
- $\overset{(4)}{\frac{dL}{d\mu_i} = \varepsilon_i = 0}$

<u>Step 3</u>: Sub (1), (2), (3) back into lagrangian function wrt $\alpha$ and $\mu$

$$L(\alpha, w) = \underbrace{\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j x_i x_j}_{\|w\|_2^2} + \frac{c}{2}\sum_{i=1}^{n}\varepsilon_i^2 - \sum_{i=1}^{n}\mu_i\varepsilon_i - \underbrace{\sum_{i=1}^{n}\alpha_i y_i w^T x_i}_{\sum \alpha_i y_i x_i}$$

$$- \underbrace{b\sum_{i=1}^{n}\alpha_i y_i}_{0} + \sum_{i=1}^{n}\alpha_i - \sum_{i=1}^{n}\alpha_i\varepsilon_i$$

$$= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j x_i x_j + \frac{c}{2}\sum_{i=1}^{n}\varepsilon_i^2 - \sum_{i=1}^{n}\mu_i\varepsilon_i - \sum_{i=1}^{n}\alpha_i\varepsilon_i - \sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j x_i^T x_j$$

$$+ \sum_{i=1}^{n}\alpha_i$$

$$= \sum_{i=1}^{n} \alpha_i + \frac{c}{2}\sum_{i=1}^{n} \varepsilon_i^2 - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i\alpha_j y_i y_j x_i^T x_j - \sum_{i=1}^{n} \mu_i \varepsilon_i - \sum_{i=1}^{n} \alpha_i \varepsilon_i$$

$\hookrightarrow$ Re-write :

$$= \underbrace{\sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i\alpha_j y_i y_j x_i^T x_j}_{\Delta} + \underbrace{\frac{c}{2}\sum_{i=1}^{n} \varepsilon_i^2 - \sum_{i=1}^{n} \mu_i \varepsilon_i - \sum_{i=1}^{n} \alpha_i \varepsilon_i}_{\text{Simplify this down further}}$$

$\therefore c = \frac{\mu_i + \alpha_i}{\varepsilon_i}$ ; $\mu_i = c\varepsilon_i - \alpha_i$ (5) ; $\alpha_i = c\varepsilon_i - \mu_i$ (6) ; $\varepsilon_i = \frac{\mu_i + \alpha_i}{c}$ (7)

*Sub in $\mu_i$ to Formula:

$$= \Delta + \frac{c}{2}\sum_{i=1}^{n} \varepsilon_i^2 - \sum_{i=1}^{n} (c\varepsilon_i - \alpha_i)\varepsilon_i - \sum_{i=1}^{n} \alpha_i \varepsilon_i$$

$$= \Delta + \frac{c}{2}\sum_{i=1}^{n} \varepsilon_i^2 - \sum_{i=1}^{n} c\varepsilon_i^2 + \cancel{\sum_{i=1}^{n} \alpha_i \varepsilon_i} - \cancel{\sum_{i=1}^{n} \alpha_i \varepsilon_i}$$

$$= \Delta + \frac{c}{2}\sum_{i=1}^{n} \varepsilon_i^2 - \sum_{i=1}^{n} c\varepsilon_i^2$$

$$= \Delta + \frac{c}{2}\sum_{i=1}^{n} \varepsilon_i^2 - c\sum_{i=1}^{n} \varepsilon_i^2 \quad \therefore \frac{c}{2} - c = -\frac{1}{2c}$$

$$= \Delta - \frac{1}{2c}\sum_{i=1}^{n} \varepsilon_i^2$$ (Note: (7) and (4) $\rightarrow$ $\varepsilon_i$ goes away and $\mu_i = \alpha_i$)

$$\boxed{\therefore \arg\max \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i\alpha_j y_i y_j x_i^T x_j - \frac{1}{2c}\sum_{i=1}^{n} \alpha_i^2}$$

# Practice

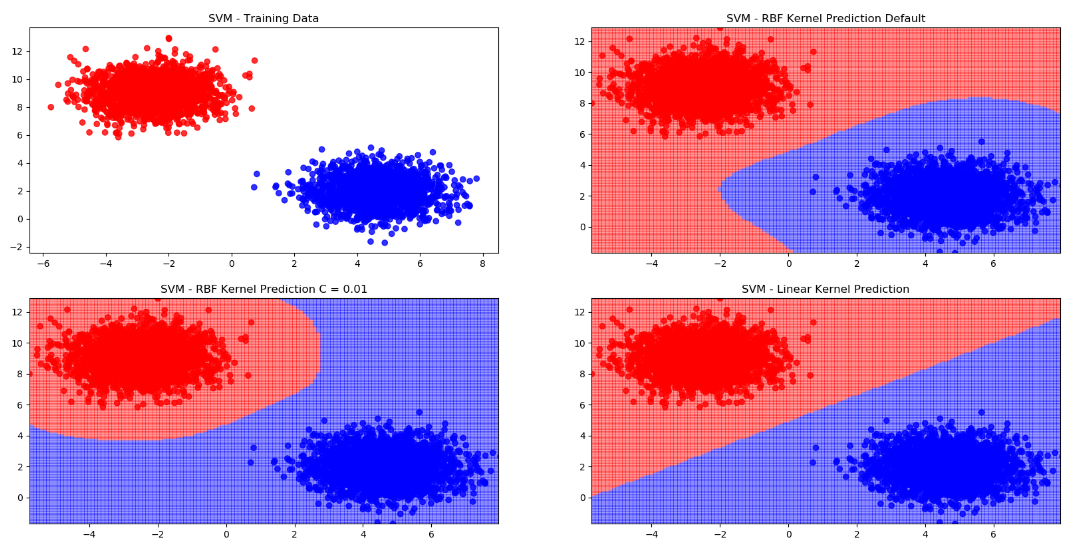
## Multilayer Perception

Output-1:

| | Classification Training Error | Classification Testing Error |
|---|---|---|
| 50HLN + no regularization + 0.01 learning rate | 0.0642592832 | 0.0195999742 |
| 250HLN + no regularization + 0.01 learning rate | 0.0716334507 | 0.0236999989 |
| 250HLN + L2 regularization + 0.01 learning rate | 0.0979647338 | 0.0303000212 |
| 50HLN + L2 regularization + 0.01 learning rate | 0.1326132715 | 0.0404000282 |

## Support Vector Machines

Output-1:




Output-2:

Error of the RBF Kernel w/ default settings = 0.00
Error of the RBF Kernel w/ C = 0.01 = 0.00
Error of the Linear SVM Kernel = 0.00