1 Dual Formulation of Soft-Margin SVM's

$$\min_{w,b,\varepsilon} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\varepsilon_i \quad \text{s.t.} \quad \forall_{i=1}^{N}: y_i\cdot(w^T\cdot\phi(x_i)+b) \geq 1-\varepsilon_i \;,\; \varepsilon_i \geq 0 \;,\; w\in\mathbb{R}^d,\; b\in\mathbb{R}$$

$\|\cdot\|$: euclidean norm ; ϕ feature map , $x_i \in \mathbb{R}^d$, $y_i \in \{-1,1\}$ labeled datapoints

Once hard-margin SVM is learned it can predict x : $f(x) = \text{sign}(w^T\phi(x)+b)$

a) State the conditions under which the solution can be obtained using Lagrange dual formulation (verify Slater's conditions)

$\lambda_i = 1-\varepsilon_i \mp \mu, \mu \geq 0$

Since $\varepsilon_i \geq 0, \forall i$ $y_i(w^T\phi(x_i)+b) \geq 1-\varepsilon_i > \lambda_i$ and therefore $-y_i(w^T\phi(x_i)+b)+\lambda_i < 0$ $\forall i \Rightarrow$ sufficient condition for strong duality according to Slater's condition $\Rightarrow$ NO CONDITIONS

b) Derive the Lagrange dual, show that it reduces to a constrained quadratic optimization problem, and state its objective function and constraints

$$\max_{\alpha,\beta\geq 0}\min_{w,b,\varepsilon} \underbrace{\frac{1}{2}\|w\|^2 + C\sum\varepsilon_i}_{\text{minimize objective}} + \underbrace{\sum\alpha_i\left[1-\varepsilon_i - y_i(w^T\phi(x_i)+b)\right]}_{\text{punish broken constraints}} + \underbrace{\sum -\beta_i\varepsilon_i}_{\text{punish } \varepsilon_i>0}$$

$$\frac{dF}{dw} = w + \sum_i \alpha_i(-y_i\phi(x_i)) = 0 \Rightarrow w = \sum\alpha_i y_i\cdot\phi(x_i) \text{ optimal } w$$

$$\frac{dF}{db} = \sum_i \alpha_i(-y_i) = 0 \quad \text{constraint}$$

$$\frac{dF}{d\varepsilon_i} = C - \alpha_i - \beta_i = 0 \Rightarrow \alpha_i + \beta_i \leq C \quad \text{constraint}$$

$$\max_{\alpha,\beta\geq 0} \frac{1}{2}\|\sum\alpha_i y_i\phi(x_i)\|^2 + C\sum\overset{0}{\varepsilon_i} + \sum\alpha_i\left[1-\overset{0}{\varepsilon_i} - y_i(\alpha_j y_j \overset{\phi(x_i)}{\phi(x_j)} + b)\right] - \sum\beta_i\overset{0}{\varepsilon_i} =$$

$$= \max_{\alpha} \frac{1}{2}\sum\alpha_i\alpha_j\cdot y_i y_j\cdot\phi(x_i)\phi(x_j) + \sum\alpha_i - \sum\alpha_i\alpha_j y_i y_j\phi(x_i)\phi(x_j) - \overset{0}{\cancel{\sum\alpha_i y_i}}\cdot b$$

$$= \max_{\alpha} \frac{-1}{2}\sum\alpha_i\alpha_j\cdot y_i y_j\cdot\phi(x_i)\phi(x_j) + \sum\alpha_i \quad \text{s.t.} \sum\alpha_i y_i = 0,\; \underbrace{0\leq\alpha_i\leq C}$$

it could be $\alpha_i+\beta_i$ but since we don't use β_i, I used $\beta_i=0$ and simplify

c) How can the solution to the primal be obtained from the dual's?

We know that $w = \sum \alpha_i y_i \Phi(x_i)$ but $b = ?$

Using the 4th KKK condition (complementarity) we know that

$$\begin{cases} \alpha_i(1-\varepsilon_i - y_i(w^T\Phi(x_i)+b)) = 0 \\ b_i \varepsilon_i = 0 \Rightarrow \underset{\alpha_i \leq C}{(C-\alpha_i)\varepsilon_i} \cancel{\Rightarrow \varepsilon_i = 0 \text{ or } \alpha_i = C} \end{cases}$$

considering that $0 \leq \alpha_i \leq C$, and enforcing

$\cancel{\text{if } \alpha_i = 0 \Rightarrow \begin{cases} 0 = 0 \\ C\varepsilon_i = 0 \end{cases}}$ $\quad$ $\cancel{\text{if } \alpha_i = C \Rightarrow}$

$\alpha_i \neq C \Rightarrow \varepsilon_i = 0$, $\quad \alpha_i \neq 0 \Rightarrow 1 \underset{}{\overset{y_i \in \{-1,1\}}{\pm}} w^T\Phi(x_i) + b = 0$ support vector

therefore if $\alpha_i \in (0, C) \Rightarrow b = \begin{cases} 1 + w^T\Phi(x_i) & \text{if } y_i = -1 \\ 1 - w^T\Phi(x_i) & \text{if } y_i = 1 \end{cases}$

d) Kernelize the dual program and the learned decision function

- $\max\limits_{\alpha} -\frac{1}{2}\sum\limits_{i,j} \alpha_i\alpha_j \cdot y_i y_j \cdot \Phi(x_i)^T\Phi(x_j) + \sum\alpha_i = -\frac{1}{2}\sum\limits_{i,j} \alpha_i\alpha_j \cdot y_i \cdot y_j \cdot K(x_i, x_j) + \sum\alpha_i$
- $f(x) = \text{sign}(w^T\Phi(x) + b) \underset{\substack{w = \sum\alpha_i y_i \Phi(x_i) \\ b = 1 - w^T\Phi(x_{sv})}}{=} \text{sign}\left[\left(\sum_i \alpha_i y_i \cdot \Phi(x_i) \cdot \Phi(x)\right) + 1 - \sum_i \alpha_i y_i \cdot \Phi(x_i)\Phi(x_{sv})\right] =$

$$= \text{sign}\left[\left(\sum\alpha_i y_i \cdot (K(x_i, x) - K(x_i, x_{sv}))\right) + 1\right]$$

2. SVMs & Quadratic Programming

$$\min_x \frac{1}{2}x^T P x + q^T x, \quad \text{s.t. } Gx \preceq h, \; Ax = b$$

a) Express P, q, G, h, A, b in terms of exercise 1

$$\max_{\alpha} -\frac{1}{2}\sum \alpha_i\alpha_j \cdot \underbrace{y_i y_j \cdot K(x_i, x_j)}_{P_{ij}} + \sum\alpha_i = \min_{\alpha} \frac{1}{2}\sum \alpha_i\alpha_j \cdot y_i y_j \cdot K(x_i, x_j) \underbrace{- \sum\alpha_i}_{-\sum 1 \cdot \alpha_i}$$

$P_{ij} = y_i y_j \cdot K(x_i, x_j) \Rightarrow \boxed{P = yy^T \odot K(x_i, x_j)}$, $\boxed{q = -\mathbb{1}}$

s.t. $0 \leq \alpha_i \leq C \begin{cases} -\alpha_i \leq 0 \\ \alpha_i \leq C \end{cases} \Rightarrow \boxed{G = \begin{pmatrix} -I \\ I \end{pmatrix}}$, $\boxed{h = \begin{pmatrix} 0 \\ C \end{pmatrix}}$

$\sum \alpha_i \cdot y_i = 0 \Rightarrow \boxed{A = y^T}$, $\boxed{b = 0}$