Math 597 Assignment 4

Jonathan Pearce
260672004

1) a) $\min_{w,b,\varepsilon} \frac{1}{2}\|w\|^2 + C\sum_i^m \varepsilon_i^p$

subject to $y_i(w\cdot x_i + b) \geq 1 - \varepsilon_i \wedge \varepsilon_i \geq 0 \quad \forall i \in [m]$

Define the lagrangian,

$$\mathcal{L}(w, b, \varepsilon, \alpha, \beta) = \frac{1}{2}\|w\|^2 + C\sum_i^m \varepsilon_i^p - \sum_i^m \alpha_i[y_i(x_i w + b) - 1 + \varepsilon_i] - \sum_i^m \beta_i \varepsilon_i$$

Satisfying KKT conditions,

$$\nabla_w \mathcal{L} = w - \sum_i^m \alpha_i y_i x_i = 0 \implies \boxed{w = \sum_i^m \alpha_i y_i x_i} \quad (1)$$

$$\nabla_b \mathcal{L} = -\sum_i^m \alpha_i y_i = 0 \implies \boxed{\sum \alpha_i y_i = 0} \quad (2)$$

$$\nabla_\varepsilon \mathcal{L} = Cp\varepsilon_i^{p-1} - \alpha_i - \beta_i = 0 \implies \boxed{\begin{aligned}\varepsilon_i^{p-1} &= \frac{\alpha_i + \beta_i}{Cp} \\ \varepsilon_i &= \left(\frac{\alpha_i + \beta_i}{Cp}\right)^{1/p-1}\end{aligned}} \quad (3)$$

$$\forall i \quad \alpha_i[y_i(x_i\cdot w + b) - 1 + \varepsilon_i] = 0 \implies \boxed{\alpha_i = 0 \text{ or } y_i(w\cdot x + b) = 1 - \varepsilon_i} \quad (4)$$

$$\forall i \quad \beta_i \varepsilon_i = 0 \implies \boxed{\beta_i = 0 \text{ or } \varepsilon_i = 0} \quad (5)$$

<u>note</u>: (1), (2), (4) and (5) are the same as when $p = 1$

$$\implies \mathcal{L} = \sum_i^m \alpha_i - \frac{1}{2}\sum_{i,j=1}^m \alpha_i\alpha_j y_i y_j (x_i\cdot x_j) + \underbrace{\sum_{i=1}^m C\varepsilon_i^p - \alpha_i\varepsilon_i - \beta_i\varepsilon_i}_{\text{need to get rit of } \varepsilon_i \text{ using } (3)}$$

$$\sum_{i=1}^{m} c\varepsilon_i^{p} - \alpha_i \varepsilon_i - \beta_i \varepsilon_i$$

$$= \sum_{i=1}^{m} c\varepsilon\left(\frac{\alpha_i + \beta_i}{cp}\right) - \alpha_i \varepsilon_i - \beta_i \varepsilon_i \qquad \text{By } ③ \quad \varepsilon^{p-1} = \frac{\alpha_i + \beta_i}{cp}$$

$$= \sum_{i=1}^{m} \varepsilon(\alpha_i + \beta_i)\left(\frac{1}{p} - 1\right)$$

$$= \sum_{i=1}^{m} \left(\frac{\alpha_i + \beta_i}{cp}\right)^{\frac{1}{p-1}} (\alpha_i + \beta_i)\left(\frac{1}{p} - 1\right) \qquad \text{By } ③ \quad \varepsilon = \left(\frac{\alpha_i + \beta_i}{cp}\right)^{\frac{1}{p-1}}$$

$$= \sum_{i=1}^{m} \frac{(\alpha_i + \beta_i)^{p/p-1}}{(cp)^{1/p-1}} \left(\frac{1}{p} - 1\right)$$

therefore the Dual is as follows:

$$\boxed{\begin{aligned} &\max_{\alpha,\beta} \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{m} \alpha_i\alpha_j y_i y_j (x_i \cdot x_j) - \sum_{i=1}^{m} \frac{(\alpha_i + \beta_i)^{\frac{p}{p-1}}}{(cp)^{\frac{1}{p-1}}}\left(1 - \frac{1}{p}\right) \\ &\text{subject to } \sum_{i=1}^{m} \alpha_i y_i = 0, \quad \alpha_i \geq 0, \quad \beta_i \geq 0 \quad \forall i \in [m] \end{aligned}}$$

b) The general case $(p > 1)$, has a 3rd summation that needs to be considered. Now we also have to maximize over β as well as α. The first 2 summations are the same as when $p=1$.

$\underline{\underline{p = 2}}$

$$\sum_{i=1}^{m} \frac{(\alpha_i + \beta_i)^2}{cp} \cdot \left(\frac{1}{2}\right) = \frac{1}{2c} \sum_{i=1}^{m} (\alpha_i + \beta_i)^2 \quad \text{convex}$$

$$\boxed{\Rightarrow \text{case } p = 2 \text{ is still convex}}$$

2) a) let $z_i = (y_1(x_1 \cdot x_i), \dots, y_m(x_m \cdot x_i))$

problem becomes,

$$\boxed{\begin{aligned} &\min_{\alpha, \beta} \frac{1}{2}\sum_{i=1}^{m} \alpha_i^2 + C\sum_{i=1}^{m} \varepsilon_i \\ &\text{Subject to } y_i(\alpha \cdot z_i + b) \geq 1 - \varepsilon_i,\ \varepsilon_i \geq 0,\ \alpha_i \geq 0\ \forall i \in [m] \end{aligned}}$$

This equivilant to the primal SVM when $p = 1$, modulo the the non-negativity constraint on α

b) Let $\beta_i \geq 0, \delta_i \geq 0, \lambda_i \geq 0$ be lagrange variables

$$\mathcal{L}(\alpha, b, \varepsilon, \beta, \delta, \lambda) = \frac{1}{2}\sum_{i=1}^{m} \alpha_i^2 + C\sum_{i=1}^{m} \varepsilon_i - \sum_{i=1}^{m} \beta_i\left[y_i\left(\sum_{j=1}^{m} \alpha_j y_j x_i \cdot x_j + b\right) - 1 + \varepsilon_i\right]$$

$$- \sum_{i=1}^{m} \delta_i \varepsilon_i - \sum_{i=1}^{m} \lambda_i \alpha_i$$

Satisfy KKT conditions:

$$\nabla_{\alpha_i} \mathcal{L} = \alpha_i - y_i x_i \cdot \sum_{j=1}^{m} \beta_j y_j x_j - \lambda_i = 0 \Rightarrow \boxed{\alpha_i = y_i x_i \cdot \sum_{j=1}^{m} \beta_j y_j x_j + \lambda_i} \quad 1$$

$$\nabla_{b_i} \mathcal{L} = -\sum_{i=1}^{m} \beta_i y_i = 0 \Rightarrow \boxed{\sum_{i=1}^{m} \beta_i y_i = 0} \quad 2$$

$$\nabla_{\varepsilon_i} = C - \beta_i - \delta_i \quad \boxed{\Rightarrow \beta_i + \delta_i = 0} \quad 3$$

$$\forall i \quad \sum_{i=1}^{m} \beta_i\left[y_i\left(\sum_{j=1}^{m} \alpha_j y_j x_j \cdot x_i + b\right) - 1 + \varepsilon_i\right] = 0$$

$$\forall i \quad \delta_i \varepsilon_i = 0, \quad \lambda_i \alpha_i = 0$$

plugging in ① to $\mathcal{L}(\alpha, b, \varepsilon; \beta, \delta, \lambda)$ we get

$$\mathcal{L}(\alpha, b, \varepsilon, \beta, \delta, \lambda) = -\frac{1}{2}\sum_{i=1}^{m} \alpha_i^2 + \sum_{i=1}^{m} \beta_i$$

$$\sum_{i=1}^{m} \alpha_i^2 = \sum_{i=1}^{m} y_i x_i \cdot \sum_{j=1}^{m} \beta_j y_j x_j - \sum_{i=1}^{m} \lambda_i^2 \quad \text{by ① and } \lambda_i \alpha_i = 0$$

$$= \sum_{i=1}^{m} \sum_{j,k=1}^{m} \beta_j \beta_k y_j y_k (x_i y_i \cdot x_j)(x_i y_i \cdot x_k) - \sum_{i=1}^{m} \lambda_i^2$$

$$= \sum_{j,k=1}^{m} \beta_j \beta_k y_j y_k \sum_{i=1}^{m} (x_i y_i \cdot x_j)(x_i y_i \cdot x_k) - \sum_{i=1}^{m} \lambda_i^2$$

therefore the dual is as follows,

$$\boxed{\begin{aligned} &\max_{\beta, \lambda} \sum_{i=1}^{m} \beta_i - \frac{1}{2}\sum_{i,j=1}^{m} \beta_i \beta_j y_i y_j \sum_{k=1}^{m} (x_k y_k \cdot x_i)(x_k y_k \cdot x_j) - \frac{1}{2}\sum_{i=1}^{m} \lambda_i^2 \\ &\text{subject to } C \geq \beta_i \geq 0,\ \lambda_i \geq 0,\ \sum_{i=1}^{m} \beta_i y_i = 0 \quad \forall i \in [m] \end{aligned}}$$

c) with $p=1$, only ① changes from the case where $p=2$

$$\nabla_{\alpha_i} \mathcal{L} = 1 - y_i x_i \left(\sum_{j=1}^{m} \beta_j y_j x_j \right) - r_i = 0 \Rightarrow \boxed{1 = y_i x_i \cdot \sum \beta_j y_j x_j + \lambda_i}$$

①

plugging ① into $\mathcal{L}$,

$$\mathcal{L}(\alpha, b, \varepsilon, \beta, \delta, \lambda) = \sum_{i=1}^{m} \alpha_i + C \sum_{i=1}^{m} \varepsilon_i - \sum_{i=1}^{m} \alpha_i (1 - \lambda_i)$$

$$- \underbrace{\sum_{i=1}^{m} \beta_i y_i b}_{0 \cdot b = 0} + \sum_{i=1}^{m} \beta_i - \sum_{i=1}^{m} \underbrace{(\beta_i + \delta_i)}_{C} \varepsilon_i - \sum \alpha_i \lambda_i$$

$$= \sum_{i=1}^{m} \alpha_i - \sum_{i=1}^{m} \alpha_i + \sum_{i=1}^{m} \beta_i$$

$$= \sum_{i=1}^{m} \beta_i$$

therefore the dual is as follows when $p=1$

$$\boxed{\max_{\beta} \sum_{i=1}^{m} \beta_i \quad \text{subject to} \quad C \geq \beta_i \geq 0, \ \sum_{i=1}^{m} \beta_i y_i = 0}$$

3) $\mathbf{K}' = \begin{bmatrix} K'(x,x) & K'(x,y) \\ K'(y,x) & K(y,y) \end{bmatrix}$

$$= \begin{bmatrix} \frac{K(x,x)}{\alpha(x)\alpha(x)} & \frac{K(x,y)}{\alpha(x)\alpha(y)} \\ \frac{K(y,x)}{\alpha(y)\alpha(x)} & \frac{K(y,y)}{\alpha(y)\alpha(y)} \end{bmatrix}$$

Since K is PDS we know $K(x,y) = K(y,x)$

$\boxed{\Rightarrow \mathbf{K}' \text{ is Symmetric}}$

Define a vector $[\beta, \lambda] \in \mathbb{R}^2$

$$[\beta,\lambda]\,\mathbf{K}' \begin{bmatrix} \beta \\ \lambda \end{bmatrix} = \frac{\beta^2 K(x,x)}{[\alpha(x)]^2} + \frac{2\beta\lambda K(x,y)}{[\alpha(x)\alpha(y)]} + \frac{\lambda^2 K(y,y)}{[\alpha(y)]^2}$$

$$= \frac{\alpha(y)^2\beta^2 K(x,x) + 2\alpha(x)\alpha(y)\beta\lambda K(x,y) + \alpha(x)^2\lambda^2 K(y,y)}{[\alpha(x)]^2[\alpha(y)]^2}$$

$[\alpha(x)]^2[\alpha(y)]^2 \geq 0 \quad \forall x, y$

$\Rightarrow$ Just need to show numerator ≥ 0

recall K is PDS, therefore for any vector $[\hat{\beta}, \hat{\lambda}]$

$$[\hat{\beta},\hat{\lambda}]\,\mathbf{K} \begin{bmatrix} \hat{\beta} \\ \hat{\lambda} \end{bmatrix} \geq 0$$

$$\Leftrightarrow \hat{\beta}^2 K(x,x) + 2\hat{\beta}\hat{\lambda} K(x,y) + \hat{\lambda}^2 K(y,y) \geq 0$$

If we let $\hat{\beta} := \alpha(y)\beta$ and $\hat{\lambda} := \alpha(x)\lambda$ we see that the numerator is ≥ 0 as well

$\boxed{\Rightarrow K' \text{ is PDS}}$

4) a) $\mathbf{K} = \begin{bmatrix} \frac{1}{2x} & \frac{1}{x+y} \\ \frac{1}{x+y} & \frac{1}{2y} \end{bmatrix}$

Clearly $\mathbf{K}$ is Symmetric. Define a vector $[\alpha, \beta] \in \mathbb{R}^2$

$$[\alpha, \beta] \begin{bmatrix} 1/2x & 1/x+y \\ \frac{1}{x+y} & \frac{1}{2y} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \frac{\alpha^2}{2x} + \frac{\alpha\beta}{x+y} + \frac{\alpha\beta}{x+y} + \frac{\beta^2}{2y}$$

$$= \frac{2y(x+y)\alpha^2 + 2x(x+y)\beta^2 + 2x\,2y\,2\alpha\beta}{4xy(x+y)}$$

$4xy(x+y) \geq 0 \quad \forall x \in [0, \infty), y \in [0, \infty)$

$\Rightarrow$ Just need to prove numerator ≥ 0

$$2\left[xy\alpha^2 + y^2\alpha^2 + x^2\beta^2 + xy\beta^2 + 4xy\alpha\beta\right]$$

$$= 2\left[\underbrace{(\alpha y + \beta x)^2}_{\geq 0} + \underbrace{xy}_{\geq 0}\underbrace{(\alpha + \beta)^2}_{\geq 0}\right]$$

≥ 0

$$\boxed{\Rightarrow K = \frac{1}{x+y} \text{ is PDS}}$$

b) $K(x, x') = \cos \angle (x, x') = \frac{x \cdot x'}{|x||x'|}$

example 6.4 in the textbook showed $x \cdot x'$ is PDS and $|x||x'| > 0$, is just a scaling of the dot product

$$\boxed{\Rightarrow K(x, x') = \cos \angle (x, x') \text{ is PDS}}$$

5) Suppose K verifies Mercer's condition.

Assume K is not PDS

Mercer's condition requires K to be symmetric

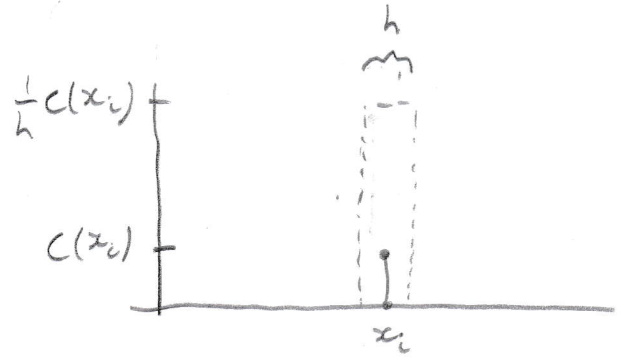
$\Rightarrow$ we are assuming $\exists\, x = (x_1, \dots x_m)$ and

$c = (c_1, \dots c_m)$ such that $\sum_{i,j=1}^{m} c_i c_j K(x_i, x_j) < 0$

let $c = (c_1, \dots c_m) = (c(x_1), \dots, c(x_m))$

Let $h \in \mathbb{R}$, $h > 0$, define $c_i^h(x)$ as follows,

$$c_i^h(x) = \begin{cases} \frac{1}{h} c(x_i) & |x - x_i| \le h/2 \\ 0 & \text{else} \end{cases}$$



note

from Mercer's condition, for any square integrable function c we have

$$\iint c(x)\, c(x')\, K(x, x')\, dx\, dx' \ge 0$$

Since $c_i^h(x)$ is square integrable we have

$$\iint c_i^h(x)\, c_j^h(x')\, k(x, x')\, dx\, dx' \geq 0$$

Since $h \geq 0$

$$\Rightarrow \lim_{h \to 0} \iint c_i^h(x)\, c_j^h(x')\, k(x, x')\, dx\, dx' \geq 0$$

<u>note</u> $\displaystyle \lim_{h \to 0} \int c_i^h(x)\, dx = \lim_{h \to 0} h \cdot \frac{1}{h} c(x_i) = c(x_i) = c_i$

$$\Rightarrow \sum_{i=1}^{m} \sum_{j=1}^{m} c_i c_j k(x_i, x_j) \geq 0 \quad ↯$$

but this contradicts our original assumption

$\Rightarrow$ K must be PDS

6) a) $\langle f, g \rangle = \int a(x)\, g(x)\, dx$

$= \int a(x)\, (q * b)(x)\, dx$

$= \int a(x) \int q(x-y)\, b(y)\, dy\, dx$ by definition of convolution

$= \int a(x) \int \Phi(y)\, b(y)\, dy\, dx$ definition of Φ

$= \iint a(x)\, b(y)\, \Phi(y)\, dy\, dx$

$= \int b(y) \int a(x)\, \Phi(x)\, dx\, dy$

$= \int b(y) \underbrace{\int a(x)\, q(y-x)\, dx}_{(q * a)(y)}\, dy$

$= \int b(y)\, (q * a)(y)\, dy$

$= \int b(y)\, f(y)\, dy$

$\boxed{\Rightarrow \text{Inner product is well defined}}$

b) $\langle \Phi(x_0,\cdot), \Phi(y_0,\cdot) \rangle$

by definition $\quad \Phi(x_0,\cdot) = \varrho(x - x_0)$

$$= (\varrho * \delta x_0)(x)$$

$f = \Phi$
$\varrho = \varrho$
$a = \delta x_0$
$\Rightarrow f = \varrho * a$

$$\Rightarrow \langle \Phi(x_0,\cdot), \Phi(y_0,\cdot) \rangle$$

$$= \langle (\varrho * \delta x_0)(x), \Phi(y_0,\cdot) \rangle$$

$$= \langle (\varrho * \delta x_0)(x), \varrho(y - y_0) \rangle$$

$$= \int \delta x_0(y)\, \varrho(y - y_0)\, dy$$

$$\langle f, g \rangle = \int a(y) g(y)\, dy$$

$$= \varrho(y - y_0) \Big|_{y = x_0}$$

$$= \varrho(x_0 - y_0)$$

$$\boxed{\Rightarrow K(x,y) = \langle \Phi(x,\cdot), \Phi(y,\cdot) \rangle}$$

c) $< h, \Phi(x,\cdot) > = \int a(x)\, \Phi(x) dx = \int h(x)\, b(x)\, dx$

$h(x) = (\varepsilon * a)(x)$

$\Phi(x) = (\varepsilon * b)(x) = (\varepsilon * \delta_{x_0})(x)$

$$\int a(x)\, \Phi(x) dx = \int a(x)\, (\varepsilon * \delta_{x_0})(x)\, dx$$

$$= \int a(x) \int \varepsilon(x-y)\, \delta_{x_0}(y)\, dy\, dx$$

$$= \int \delta_{x_0}(y) \int a(x)\, \varepsilon(x-y)\, dx\, dy$$

$$= \int \delta_{x_0}(y)\, (\varepsilon * a)(x)\, dy$$

$$= \int \delta_{x_0}(y)\, h(x)\, dy$$

$$= h(x) \underbrace{\int \delta_{x_0}(y)\, dy}_{=1}$$

$$= h(x)$$

$$\boxed{\Rightarrow h(x) = < h, \Phi(x,\cdot) >}$$