KERNEL METHODS IN M.L.

Homework 3 – by DAVID SOTO

Exercise 1. Support Vector Classifier.

We consider a dataset of N pairs $(x_i, y_i) \in \mathbb{R}^d \times \{-1, 1\}$.
We consider functions f that belong to a RKHS $\mathcal{H}$ of kernel K.
We have the following optimization problem:

$$(1) \quad \min_{f, b, \xi_i} \frac{1}{2}\|f\|^2 + C\sum_{i=1}^{N} \xi_i$$
$$\text{s.t.} \quad y_i(f(x_i) + b) \geq 1 - \xi_i \quad \forall i = 1, \dots, N$$
$$\xi_i \geq 0 \quad \forall i = 1, \dots, N$$

1. **a.** After calculation, we obtain that the Lagrangian of the problem (1) is given by:

$$L(f, b, \xi, \alpha, \mu) = \frac{1}{2}\|f\|^2 - \sum_{i=1}^{N} \alpha_i y_i f(x_i) + \sum_{i=1}^{N} \xi_i (C - \alpha_i - \mu_i) - b\sum_{i=1}^{N} \alpha_i y_i + \sum_{i=1}^{N} \alpha_i .$$

where $\alpha_i \geq 0$ are the N dual parameters corresponding the margin inequalities and $\mu_i \geq 0$ are the N dual parameters corresponding to the positivity constraints on ξ_i.

b. The lagrangian L is an affine function of ξ_i, and an affine function of b,

$\Rightarrow$ its minimum is $-\infty$ except when

$$\forall i = 1, \dots, N, \quad \alpha_i + \mu_i = C$$
and
$$\sum_{i=1}^{N} \alpha_i y_i = 0$$

Moreover, using the reproducing property, we can easily prove that

$$\frac{1}{2}\|f\|^2 - \sum_{i=1}^{N} \alpha_i y_i f(x_i) = \frac{1}{2}\|f\|^2 - \langle f, \sum_{i=1}^{N} \alpha_i y_i K_{x_i} \rangle_{\mathcal{H}}.$$

$\Rightarrow$ L is a convex function of f. Therefore, we can minimize over f by setting the gradient equal to 0, this yields:

$$\Longleftrightarrow \quad f = \sum_{i=1}^{N} \alpha_i y_i K_{x_i}$$

It follows that the dual function is given by:

$$q(\alpha, \mu) = \inf_{f, b, \xi} L(f, b, \xi, \alpha, \mu)$$

$$q(\alpha, \mu) = \begin{cases} \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \cdot K(x_i, x_j) & \text{if } \alpha + \mu = C\mathbb{1},\ \sum_{i=1}^{N} \alpha_i y_i = 0 \\ -\infty & \text{otherwise} \end{cases}$$

Hence, the dual problem is given by:

$$\boxed{\begin{aligned} \max_{\alpha, \mu} \quad & \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \\ \text{s.t.} \quad & \alpha + \mu = C\mathbb{1} \\ & \sum_{i=1}^{N} \alpha_i y_i = 0 \\ & \alpha \geq 0 \\ & \mu \geq 0 \end{aligned}}$$

We can eliminate the parameter μ, & change the dual problem into a minimization problem, ~~this yields:~~

Finally, the dual problem is given by:

$\longrightarrow$

(Dual)

$$\min_{\alpha} \quad \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j K(x_i, x_j) - \sum_{i=1}^{N} \alpha_i$$

$$\text{s.t.} \quad 0 \leq \alpha_i \leq C \quad \forall i = 1, \dots, N$$

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

Moreover, we have that $\forall x$:

$$f(x) = \sum_{i=1}^{N} \alpha_i y_i K(x_i, x)$$

c. We can rewrite the dual as follows:

(Dual)

$$\min_{\alpha} \quad \frac{1}{2} \alpha^T \operatorname{diag}(y)\, K \operatorname{diag}(y)\, \alpha - \alpha^T 1$$

$$\text{s.t.} \quad 0 \leq \alpha \leq C 1$$

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

This is the minimization of a * convex quadratic function with affine constraints $\Rightarrow$ Slater's condition is satisfied $\Rightarrow$ Strong duality holds!

* because K is a positive definite kernel! and the function $[\alpha \longmapsto -\alpha^T 1]$ is affine.

d) The complementary slackness conditions are $\forall i = 1, \dots, N$.

$$\begin{cases} \alpha_i \left(y_i \left(f(x_i) + b \right) + \xi_i - 1 \right) = 0 \\ \mu_i \, \xi_i = 0 \end{cases}$$

if $\alpha_i > 0 \Rightarrow y_i \left(f(x_i) + b \right) + \xi_i - 1 = 0$

$\quad$ if $\xi_i > 0 \Rightarrow y_i \left(f(x_i) + b \right) < 1$

$\quad$ if $\xi_i = 0 \Rightarrow y_i \left(f(x_i) + b \right) = 1$.

Exercise 2 : Kernel Support Vector Regression.

1. We consider a dataset of N pairs $(x_i, y_i) \in \mathbb{R}^d \times \mathbb{R}$.
We consider functions f that belong to a RKHS $\mathcal{H}$ of kernel K.
We have the following optimization problem:

$$\text{(2)} \quad \begin{aligned} \min_{f, b, \xi^+, \xi^-} \quad & \frac{1}{2} \|f\|^2 + C \sum_{i=1}^{N} [\xi_i^+ + \xi_i^-] \\ \text{s.t.} \quad & y_i - f(x_i) - b \leq \eta + \xi_i^+ \\ & -y_i + f(x_i) + b \leq \eta + \xi_i^- \\ & \xi_i^+ \geq 0 \\ & \xi_i^- \geq 0 \end{aligned}$$

a. After calculation, we obtain that the Lagrangian of the problem (2) is given by:

$$\begin{aligned} L(f, b, \xi^+, \xi^-, \alpha^+, \alpha^-, \mu^+, \mu^-) = & \frac{1}{2} \|f\|^2 + \sum_{i=1}^{N} f(x_i)(\alpha_i^- - \alpha_i^+) + b \sum_{i=1}^{N} (\alpha_i^- - \alpha_i^+) \\ & + \sum_{i=1}^{N} \xi_i^+ (C - \alpha_i^+ - \mu_i^+) + \sum_{i=1}^{N} \xi_i^- (C - \alpha_i^- - \mu_i^-) \\ & + \sum_{i=1}^{N} \alpha_i^+ (y_i - \eta) - \sum_{i=1}^{N} \alpha_i^- (-y_i - \eta) \end{aligned}$$

where $(\alpha_i^+)_i \geq 0$ and $(\alpha_i^-)_i \geq 0$ are the $2N$ dual parameters corresponding to the tube inequalities, and where $(\mu_i^+)_i$ and $(\mu_i^-)_i$ are the N dual parameters corresponding to the positivity constraints on ξ_i^+ and ξ_i^-.

6. $\forall i = 1, \dots, N$, the lagrangian L is an affine function of ξ_i^+,
$\forall i = 1, \dots, N$, the lagrangian L is an affine function of ξ_i^-
and the lagrangian L is an affine function of b.

$\Rightarrow$ its minimum is $-\infty$ except, (at least) when
$$\sum_{i=1}^{N} (\alpha_i^- - \alpha_i^+) = 0$$

$\forall i = 1, \dots, N, \quad \alpha_i^+ + \mu_i^+ = C$
$\forall i = 1, \dots, N, \quad \alpha_i^- + \mu_i^- = C$

Moreover, using the reproducing property we have that :

$$\frac{1}{2}\|f\|^2 - \sum_{i=1}^{N} f(x_i)(\alpha_i^+ - \alpha_i^-) = \frac{1}{2}\|f\|^2 - \langle f, \sum_{i=1}^{N} (\alpha_i^+ - \alpha_i^-) K_{x_i} \rangle_{\mathcal{H}}$$

With this result, we can easily see that L is a convex function of f. Therefore we minimize over f by setting the gradient with respect to f equal to 0.
This yields:

$$f = \sum_{i=1}^{N} (\alpha_i^+ - \alpha_i^-) K_{x_i}$$

It follows that the dual problem is given by:

$$(\text{Dual}) \quad \max_{\alpha^+,\alpha^-,\mu^+,\mu^-} \; -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\alpha_i^+-\alpha_i^-)(\alpha_j^+-\alpha_j^-)K(x_i,x_j)+\sum_{i=1}^{N}\alpha_i^+(y_i-\eta)-\sum_{i=1}^{N}\alpha_i^-(y_i-\eta)$$

$$\text{s.t.} \quad \sum_{i=1}^{N}(\alpha_i^- - \alpha_i^+) = 0$$

$$\forall i=1,\dots,N, \quad \alpha_i^+ + \mu_i^+ = C$$

$$\forall i=1,\dots,N, \quad \alpha_i^- + \mu_i^- = C$$

$$\forall i=1,\dots,N, \quad \alpha_i^+, \alpha_i^-, \mu_i^+, \mu_i^- \geq 0$$

which is equivalent to the following:

$$(\text{Dual}) \quad \min_{\alpha^+,\alpha^-} \; \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\alpha_i^+-\alpha_i^-)(\alpha_j^+-\alpha_j^-)K(x_i,x_j)-\sum_{i=1}^{N}\alpha_i^+(y_i-\eta) + \sum_{i=1}^{N}\alpha_i^-(-y_i-\eta)$$

$$\text{s.t.} \quad \sum_{i=1}^{N}(\alpha_i^- - \alpha_i^+) = 0$$

$$0 \leq \alpha_i^+ \leq C \quad \forall i=1,\dots,N$$

$$0 \leq \alpha_i^- \leq C \quad \forall i=1,\dots,N$$

Moreover we have that $\forall x$,

$$f(x) = \sum_{i=1}^{N}(\alpha_i^+ - \alpha_i^-)\,K(x_i,x)$$