[1] In our derivation of the support vector machine (SVM) model, the sample points with labels $(x^{(1)}, y_1) \cdots (x^{(n)}, y_n)$ are assumed to be linearly separable. The resulting model is known as the hard margin SVM. When the data is not linearly separable, the following model was proposed by Vapnik, known as the soft margin SVM:

$$\min_{\omega, b, \xi_1, \cdots, \xi_n} \sum_{i=1}^{n} \xi_i + \frac{1}{2}\|\omega\|^2$$

subject to $\xi_i \geq 0$ ; $y_i(\omega^T x^{(i)} + b) \geq 1 - \xi_i \quad i = 1, \cdots, n$

Here, $\xi_1, \cdots, \xi_n$ are refered to as slack variables. For a variable $x \in \mathbb{R}$, we introduce the notation

$$x_+ := \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

Prove that the above minimization problem is equivalent to the following regularized risk form:

$$\min_{\omega, b} \sum_{i=1}^{n} [1 - y_i(\omega^T x^{(i)} + b)]_+ + \frac{1}{2}\|\omega\|^2$$

Proof

From the constraints on $\xi_i$ we have:

$$\xi_i \geq 1 - y_i(\omega^T x^{(i)} + b)$$

and

$$\xi_i \geq 0$$

Thus,

$$\min_{\xi_i} \xi_i = \begin{cases} 1 - y_i(\omega^T x^{(i)} + b) & \text{if } 1 - y_i(\omega^T x^{(i)} + b) \geq 0 \\ 0 & \text{if } 1 - y_i(\omega^T x^{(i)} + b) < 0 \end{cases}$$

Which is equivalent to

$$(*) \quad \min_{\xi_i} \xi_i = (1 + y_i(\omega^T x^{(i)} + b))_+$$

Next Page

$\boxed{1 \text{ cont.}}$ Notice now, since $\xi_i$ are $n$ independent variables, i.e. $\xi_i$ does NOT depend on $\xi_j$ for $\forall i \neq j = 1, \cdots, n$, we have

$$\min_{\xi_1, \cdots, \xi_n} \sum_{i=1}^{n} \xi_i = \sum_{i=1}^{n} \min_{\xi_i} \xi_i$$

Thus,

$$\min_{\xi_1, \cdots, \xi_n} \sum_{i=1}^{n} \xi_i = \sum_{i=1}^{n} \left[1 + y_i(\omega^T x^{(i)} + b)\right]_+ \quad \text{from } (*)$$

Finally, since $\frac{1}{2}\|\omega\|^2$ is constant with respect to $\xi_1, \cdots, \xi_n$ we have

$$\min_{\omega, b, \xi_1, \cdots, \xi_n} \sum_{i=1}^{n} \xi_i + \frac{1}{2}\|\omega\|^2 = \min_{\omega, b}\left[\left(\min_{\xi_1, \cdots, \xi_n} \sum_{i=1}^{n} \xi_i\right) + \frac{1}{2}\|\omega\|^2\right]$$

Which shows

$$\min_{\omega, b, \xi_1, \cdots, \xi_n} \sum_{i=1}^{n} \xi_i + \frac{1}{2}\|\omega\|^2 = \min_{\omega, b} \sum_{i=1}^{n} \left[1 + y_i(\omega^T x^{(i)} + b)\right]_+ + \frac{1}{2}\|\omega\|^2$$

■

[2] Let $f, h: \mathbb{R}^n \to \mathbb{R}$ be real-valued functions. Recall that a necessary condition for $x^* \in \mathbb{R}^n$ to be a local maximizer of the following constraint optimization problem

$$\max f(x)$$
$$\text{subject to } h(x) = 0$$

is $\frac{\partial L}{\partial x}(x^*) = 0$, we $L$ is the Lagrange multiplier $L(x,\lambda) := f(x) - \lambda h(x)$.

Given two matrices $A, B \in \mathbb{R}^{n\times n}$, a complex number $\lambda \in \mathbb{C}$ is called an <u>eigenvalue of the pair $(A,B)$</u> if there exists a non-zero vector $v \in \mathbb{C}^n$, known as the <u>associated eigenvector of the pair $(A,B)$</u>, such that $Av = \lambda B v$.

[2a] Let $A, B$ be two symmetric matrices. If $x$ is an eigenvector of the pair $(A,B)$ associated to the eigenvalue $\lambda$ and $\bar{x}^T B x \neq 0$ where $\bar{x}$ denotes the complex conjugate of $x$, prove that $\lambda$ is a real number.

<u>Proof</u>

By definition, $Ax = \lambda B x$ where $A, B \in \mathbb{R}^{n\times n}$ are symmetric matrices and $x \in \mathbb{C}^n$ and $\lambda \in \mathbb{C}$.

Multiplying from the left by $\bar{x}^T$ yields:

$$\bar{x}^T A x = \lambda \bar{x}^T B x$$

Since $\bar{x}^T B x \neq 0$ we have:

$$\lambda = \frac{\bar{x}^T A x}{\bar{x}^T B x} \quad (\star)$$

Now, let $S \in \mathbb{R}^{n\times n}$ be any symmetric $n\times n$ matrix and $v \in \mathbb{C}^n$ be any complex size-$n$ column vector.

Notice then:

$$(\bar{v}^T S v)^T = v^T S^T (\bar{v}^T)^T = v^T S \bar{v} \quad \text{since } S \text{ is symmetric}$$

Moreover, since $\bar{v}^T S v$ is a scalar, $(\bar{v}^T S v)^T = \bar{v}^T S v$, so:

$$\bar{v}^T S v = v^T S \bar{v}$$

Now we recall some basic rules of complex conjugation for scalars.

Next Page

$\boxed{2a \text{ cont.}}$ For any $a, b \in \mathbb{C}$:

$$\overline{ab} = \bar{a}\bar{b}, \quad \overline{a+b} = \bar{a} + \bar{b}, \quad \bar{\bar{a}} = a$$

And, iff $\bar{a} = a$, then $a \in \mathbb{R}$.

These rules generalize to vectors and matrices as well:

Let $a, b \in \mathbb{C}^n$ and $A, B \in \mathbb{C}^{n \times n}$ then:

$$\overline{a^T b} = \overline{\sum_i a_i b_i} = \sum_i \bar{a}_i \bar{b}_i = \bar{a}^T \bar{b} \quad \text{and } (\bar{\bar{a}})_i = \bar{\bar{a}}_i = a_i = (a)_i$$

$$(\overline{Ab})_i = \overline{\sum_j A_{ij} b_j} = \sum_j \bar{A}_{ij} \bar{b}_j = (\bar{A}\bar{b})_i$$

$$(\overline{AB})_{ij} = \overline{\sum_k A_{ik} B_{kj}} = \sum_k \bar{A}_{ik} \bar{B}_{kj} = (\bar{A}\bar{B})_{ij}$$

Using these rules, we can calculate $\overline{\bar{v}^T S v}$:

$$\overline{\bar{v}^T S v} = \bar{\bar{v}}^T \bar{S} \bar{v} = v^T S \bar{v} \quad \text{since } S \in \mathbb{R}^{n \times n}$$

Thus, from before we have:

$$\overline{\bar{v}^T S v} = v^T S \bar{v} = \bar{v}^T S v$$

Therefore $\bar{v}^T S v \in \mathbb{R}$. Applying this to $\bar{x}^T A x$ and $\bar{x}^T B x$ in ⊛, we have

$$\bar{x}^T A x \in \mathbb{R} \quad \text{and } \bar{x}^T B x \neq 0 \in \mathbb{R}$$

Which proves $\lambda \in \mathbb{R}$.

■

[2b] Let $A, B$ be two symmetric matrices. If $x^* \in \mathbb{R}^n$ is a local maximizer of the following optimization problem

$$\max x^T A x$$
$$\text{subject to } x^T B x = 1$$

prove that $x^*$ is an eigenvector of the pair $(A, B)$.

Proof

In this problem, $f(x) = x^T A x$

$$h(x) = x^T B x - 1 = 0$$

So the Lagrange multiplyer is:

$$L(x, \lambda) = x^T A x - \lambda (x^T B x - 1)$$

Since $x^*$ is a local maximizer, we know

$$\frac{\partial L}{\partial x}(x^*) = 0$$

So:

$$2Ax^* - \lambda \cdot 2Bx^* = 0$$

$$\Rightarrow \quad Ax^* = \lambda B x^*$$

Thus $x^*$ is an eigenvector of the pair $(A, B)$. ∎

$\boxed{13}$ Let $A, B \in \mathbb{R}^{n \times n}$ be two symmetric matrices, and

$$R(x) := \frac{x^T A x}{x^T B x}$$

be the generalized Rayleigh quotient. If $x^*$ is a critical point of $R(x)$, prove that $x^*$ is an eigenvector of the pair $(A, B)$ associated to the eigenvalue $R(x^*)$ of the pair $(A, B)$.

Proof

Since $x^*$ is a critical point of $R(x)$, we know

$$\frac{\partial R}{\partial x}(x^*) = 0$$

So

$$\frac{(x^{*T} B x^*) \cdot 2Ax^* - (x^{*T} A x^*) \cdot 2Bx^*}{((x^*)^T B x^*)^2} = 0 \quad \left(\begin{array}{c}\text{quotient}\\ \text{rule}\end{array}\right)$$

$$\Longrightarrow (x^{*T} B x^*) A x^* - (x^{*T} A x^*) B x^* = 0$$

Now since $x^{*T} B x^*$ is a scalar, we divide by it and rearrange to get:

$$Ax^* = \left(\frac{x^{*T} A x^*}{x^{*T} B x^*}\right) B x^*$$

Since $\frac{x^{*T} A x^*}{x^{*T} B x^*} = R(x^*)$ (which is a scalar) we have shown

$$Ax^* = R(x^*) B x^*$$

Which proves that $x^*$ is an eigenvector of the pair $(A, B)$ with eigenvalue $R(x^*)$ of the pair $(A, B)$.

■