4 Lagrangian Duality

Consider the following nonlinear optimization problem:

- **Primal Problem (P)**

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & \mathbf{g}(x) \leq \mathbf{0}, \mathbf{h}(x) = \mathbf{0} \\ \text{variables} & x \in \mathcal{X} \end{array}$$

- **Lagrangian Dual Problem (D)**

$$\begin{array}{ll} \text{maximize} & \theta(\mathbf{u}, \mathbf{v}) = \inf_{x \in \mathcal{X}} \overbrace{\left(f(x) + \mathbf{u}^T \mathbf{g}(x) + \mathbf{v}^T \mathbf{h}(x)\right)}^{L(x, \mathbf{u}, \mathbf{v})} \\ \text{subject to} & \mathbf{u} \geq \mathbf{0} \\ \text{variables} & \mathbf{u} \in \mathbb{R}^m, \mathbf{v} \in \mathbb{R}^\ell \end{array} \tag{2}$$

  where L is referred as the **Lagranian function**.

4.1 Duality Theorem

Theorem 4.1 (Weak Duality Theorem)**.** *Let x be a feasible solution to Problem P, and let $(\mathbf{u}, \mathbf{v})$ be a feasible solution to Problem D. Then*

$$\theta(\mathbf{u}, \mathbf{v}) = \inf_{\tilde{x} \in \mathcal{X}} L(\tilde{x}, \mathbf{u}, \mathbf{v}) \leq L(x, \mathbf{u}, \mathbf{v}) = f(x) + \mathbf{u}^T \mathbf{g}(x) + \mathbf{v}^T \mathbf{h}(x) \leq f(x)$$

Corollary 4.2.

$$\sup\{\theta(\mathbf{u}, \mathbf{v}) : \mathbf{u} \geq \mathbf{0}, \mathbf{v} \in \mathbb{R}^\ell\} \leq \inf\{f(x) : x \in \mathcal{X}, \mathbf{g}(x) \leq \mathbf{0}, \mathbf{h}(x) = \mathbf{0}\}.$$

Lemma 4.3 (Supporting Hyperplane Theorem)**.** *Let S be a convex set in $\mathbb{R}^n$ and $\mathbf{x}_0 \notin S^\circ$. Then there exists nonzero vector $\mathbf{w} \in \mathbb{R}^n$ such that $\mathbf{w}^T(\mathbf{x} - \mathbf{x}_0) \leq 0$ for all $\mathbf{x} \in \bar{S}$.*

Proof. See Functional Analysis Exercise 3.1. □

Lemma 4.4. *Let $\mathcal{X}$ be a convex set in $\mathbb{R}^n$. Let $\alpha : \mathbb{R}^n \to \mathbb{R}$ and $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^m$ be (componentwise) convex, and $\mathbf{h} : \mathbb{R}^n \to \mathbb{R}^\ell$ be affine (that is, $\mathbf{h}$ is of the form $\mathbf{h}(\mathbf{x}) = \mathbf{A}\mathbf{x} - \mathbf{b}$). Also, let u_0 be a scalar, $\mathbf{u} \in \mathbb{R}^m$ and $\mathbf{v} \in \mathbb{R}^\ell$. Consider the following two systems:*

> **System 1:** $\alpha(x) < 0$, $\mathbf{g}(\mathbf{x}) \leq \mathbf{0}$, $\mathbf{h}(\mathbf{x}) = \mathbf{0}$ *for some* $\mathbf{x} \in \mathcal{X}$.
>
> **System 2:** $u_0 \alpha(x) + \mathbf{u}^T \mathbf{g}(\mathbf{x}) + \mathbf{v}^T \mathbf{h}(\mathbf{x}) \geq 0$ *for some* $(u_0, \mathbf{u}, \mathbf{v}) \neq (0, \mathbf{0}, \mathbf{0})$, $(u_0, \mathbf{u}) \geq (0, \mathbf{0})$ *and for all* $\mathbf{x} \in \mathcal{X}$.

If System 1 has no solution $\mathbf{x}$, then System 2 has a solution $(u_0, \mathbf{u}, \mathbf{v})$. Conversely, if System 2 has a solution $(u_0, \mathbf{u}, \mathbf{v})$ with $u_0 > 0$, then System 1 has no solution.

Proof.

- Assume that System 1 has no solution. Define convex set

$$S = \{(p, \mathbf{q}, \mathbf{r}) : p > \alpha(\mathbf{x}), \mathbf{q} \geq \mathbf{g}(\mathbf{x}), \mathbf{r} = \mathbf{h}(\mathbf{x}) \text{ for some } \mathbf{x} \in \mathcal{X}\}$$

  Then $(0, \mathbf{0}, \mathbf{0}) \notin S$, and by supporting hyperplane theorem there exists $(u_0, \mathbf{u}, \mathbf{v}) \neq (0, \mathbf{0}, \mathbf{0})$ such that

$$u_0 p + \mathbf{u}^T \mathbf{q} + \mathbf{v}^T \mathbf{r} \geq 0, \quad \forall (p, \mathbf{q}, \mathbf{r}) \in \bar{S}$$

  Since $(p, \mathbf{q}, \mathbf{r}) \in S$ implies $(\tilde{p}, \tilde{\mathbf{q}}, \mathbf{r}) \in S$ for all $\tilde{p} \geq p$, $\tilde{\mathbf{q}} \geq \mathbf{q}$, one must have $u_0 \geq 0$ and $\mathbf{u} \geq \mathbf{0}$. For arbitrary $\mathbf{x} \in \mathcal{X}$, note that $(\alpha(\mathbf{x}), \mathbf{g}(\mathbf{x}), \mathbf{h}(\mathbf{x})) \in \bar{S}$, therefore

$$u_0 \alpha(\mathbf{x}) + \mathbf{u}^T \mathbf{g}(\mathbf{x}) + \mathbf{v}^T \mathbf{h}(\mathbf{x}) \geq 0.$$

  One concludes that System 2 has a solution.

- Assume that System 2 has a solution $(u_0, \mathbf{u}, \mathbf{v})$ with $u_0 > 0$. For any $\mathbf{x} \in \mathcal{X}$ such that $\mathbf{g}(\mathbf{x}) \leq \mathbf{0}$, $\mathbf{h}(\mathbf{x}) = \mathbf{0}$, one must have $\alpha(\mathbf{x}) \geq 0$. Hence System 1 has no solution.

□

Theorem 4.5 (Strong Duality Theorem)**.** *Let $\mathcal{X}$ be a nonempty convex set in $\mathbb{R}^n$. Let $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ be convex, and $h : \mathbb{R}^n \to \mathbb{R}^\ell$ be affine. If there exists $\hat{\mathbf{x}} \in \mathcal{X}$ such that $\mathbf{g}(\hat{\mathbf{x}}) < \mathbf{0}$ and $\mathbf{h}(\hat{\mathbf{x}}) = \mathbf{0}$ and $\mathbf{0} \in \mathbf{h}(\mathcal{X})^{\circ}$, then*

$$\underbrace{\inf\{f(\mathbf{x}) : \mathbf{x} \in \mathcal{X}, \mathbf{g}(\mathbf{x}) \leq \mathbf{0}, \mathbf{h}(\mathbf{x}) = \mathbf{0}\}}_{\textit{defined as } \gamma} = \underbrace{\sup\{\theta(\mathbf{u}, \mathbf{v}) : \mathbf{u} \geq \mathbf{0}, \mathbf{v} \in \mathbb{R}^\ell\}}_{\textit{defined as } \zeta} \quad (3)$$

where $\theta(\mathbf{u}, \mathbf{v})$ is defined in (2). Furthermore, if γ is finite, then (1) there exists optimal dual feasible solution $(\bar{\mathbf{u}}, \bar{\mathbf{v}})$ such that $\theta(\bar{\mathbf{u}}, \bar{\mathbf{v}}) = \zeta$; (2) If there exists optimal primal feasible solution $\bar{\mathbf{x}}$ such that $f(\bar{\mathbf{x}}) = \gamma$, then $\bar{\mathbf{u}}^T \mathbf{g}(\bar{\mathbf{x}}) = 0$.

Proof. Note that $\gamma < \infty$ since $\hat{\mathbf{x}}$ is a primal feasible solution. If $\gamma = -\infty$, then Corollary.4.2 implies $\zeta = -\infty$ and hence (3) is satisified.

Thus, suppose that γ is finite. Since the following system

$$f(\mathbf{x}) - \gamma < 0, \; \mathbf{g}(\mathbf{x}) \leq \mathbf{0}, \; \mathbf{h}(\mathbf{x}) = \mathbf{0}, \text{ for some } \mathbf{x} \in \mathcal{X}$$

has no solution, the previous lemma implies there exists $(u_0, \mathbf{u}, \mathbf{v}) \neq (0, \mathbf{0}, \mathbf{0})$, $(u_0, \mathbf{u}) \geq (0, \mathbf{0})$ such that

$$u_0[f(\mathbf{x}) - \gamma] + \mathbf{u}^T \mathbf{g}(\mathbf{x}) + \mathbf{v}^T \mathbf{h}(\mathbf{x}) \geq 0, \forall \mathbf{x} \in \mathcal{X} \quad (4)$$

- *Claim:* $u_0 > 0$

  Suppose $u_0 = 0$, then (4) becomes

$$\mathbf{u}^T \mathbf{g}(\mathbf{x}) + \mathbf{v}^T \mathbf{h}(\mathbf{x}) \geq 0, \forall \mathbf{x} \in \mathcal{X}$$

  Take $\mathbf{x} = \hat{\mathbf{x}}$ implies $\mathbf{u} = \mathbf{0}$. Since $\mathbf{0} \in \mathbf{h}(\mathcal{X})^{\circ}$, there exists $\tilde{\mathbf{x}} \in \mathcal{X}$ such that $\mathbf{h}(\tilde{\mathbf{x}}) = -\lambda \mathbf{v}$ for some $\lambda > 0$. Take $\mathbf{x} = \tilde{\mathbf{x}}$ implies $\mathbf{v} = \mathbf{0}$. Therefore $(u_0, \mathbf{u}, \mathbf{v}) = (0, \mathbf{0}, \mathbf{0})$, which is a contradiction.

Take $\bar{\mathbf{u}} = \frac{\mathbf{u}}{u_0}$ and $\bar{\mathbf{v}} = \frac{\mathbf{v}}{u_0}$, then by (4) one obtains

$$f(\mathbf{x}) + \bar{\mathbf{u}}^T \mathbf{g}(\mathbf{x}) + \bar{\mathbf{v}}^T \mathbf{h}(\mathbf{x}) \geq \gamma, \ \forall \mathbf{x} \in \mathcal{X} \tag{5}$$

Therefore $\theta(\bar{\mathbf{u}}, \bar{\mathbf{v}}) \geq \gamma$. Since $\theta(\bar{\mathbf{u}}, \bar{\mathbf{v}}) \leq \zeta \leq \gamma$ by Corollary.4.2, one concludes that $\gamma = \zeta$ and that $(\bar{\mathbf{u}}, \bar{\mathbf{v}})$ solves the dual problem.

Finally, assume $\bar{\mathbf{x}}$ is primal feasible such that $f(\bar{\mathbf{x}}) = \gamma$, then substitute $\mathbf{x} = \bar{\mathbf{x}}$ into (5) gives $\bar{\mathbf{u}}^T \mathbf{g}(\bar{\mathbf{x}}) = 0$. □

4.2 KKT Conditions

Theorem 4.6. *Let x and $(\mathbf{u}, \mathbf{v})$ be primal and dual feasible solutions respectively. Then $f(x) = \theta(\mathbf{u}, \mathbf{v})$ iff the* **KKT conditions** *hold:*

- *Stationary Condition:* $x \in \arg\min_{\tilde{x} \in \mathcal{X}} L(\tilde{x}, \mathbf{u}, \mathbf{v})$.
- *Complementary Slackness:* $u_i g_i(x) = 0$ *for all* $i = 1, ..., n$.
- *Primal Feasibility:* $\mathbf{g}(x) \leq \mathbf{0}, \mathbf{h}(x) = \mathbf{0}$
- *Dual Feasibility:* $\mathbf{u} \geq \mathbf{0}$.

Proof.

- *Necessity:* Suppose $\bar{x}$ and $\bar{\mathbf{u}}, \bar{\mathbf{v}}$ are primal and dual feasible solutions respectively for which $f(\bar{x}) = \theta(\bar{\mathbf{u}}, \bar{\mathbf{v}})$. The feasibility conditions follow by definition. Recall Theorem.4.1, one has

  $$f(\bar{x}) = \theta(\bar{\mathbf{u}}, \bar{\mathbf{v}}) = \inf_{x \in \mathcal{X}} L(x, \bar{\mathbf{u}}, \bar{\mathbf{v}}) \leq L(\bar{x}, \bar{\mathbf{u}}, \bar{\mathbf{v}}) = f(\bar{x}) + \bar{\mathbf{u}}^T \mathbf{g}(\bar{x}) + \bar{\mathbf{v}}^T \mathbf{h}(\bar{x}) \leq f(\bar{x})$$

  Hence we can replace the inequalities with equalities. Note that $\inf_{x \in \mathcal{X}} L(x, \bar{\mathbf{u}}, \bar{\mathbf{v}}) = L(\bar{x}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$, so the stationary condition holds. Note that $\mathbf{h}(\bar{x}) = \mathbf{0}$, so $\bar{\mathbf{u}}^T \mathbf{g}(\bar{x}) = 0$, which implies $\bar{u}_i g_i(\bar{\mathbf{x}}) = 0$ for all $i = 1, ..., n$, so the complementary slackness holds.

- *Sufficiency:* Suppose $\bar{x}, \bar{\mathbf{u}}, \bar{\mathbf{v}}$ satisify KKT conditions, then

  $$\theta(\bar{\mathbf{u}}, \bar{\mathbf{v}}) = f(\bar{x}) + \bar{\mathbf{u}}^T \mathbf{g}(\bar{x}) + \bar{\mathbf{v}}^T \mathbf{h}(\bar{x}) = f(\bar{x})$$

  where the first equality comes from stationary condition, and the second equality comes from complementary slackness and primal feasibility.

□

5 Support Vector Machines

5.1 Support Vector Classification (SVC)

The SVC optimization problem:

$$\begin{array}{ll} \text{minimize} & f(\mathbf{w}, b, \boldsymbol{\xi}) = \frac{1}{2}\|\mathbf{w}\|^2 + \sum_{i=1}^{N} C_i \xi_i \\ \text{subject to} & \begin{array}{l} y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 - \xi_i \\ \xi_i \geq 0 \end{array} , \ i = 1, ..., N \\ \text{variables} & \mathbf{w} \in \mathbb{R}^m, b \in \mathbb{R}, \boldsymbol{\xi} \in \mathbb{R}^N \end{array}$$

Rewrite it in the form of **primal problem**:

$$\begin{array}{ll} \text{minimize} & f(\mathbf{w}, b, \boldsymbol{\xi}) = \frac{1}{2}\|\mathbf{w}\|^2 + \sum_{i=1}^{N} C_i \xi_i \\ \text{subject to} & \begin{array}{l} g_{1,i}(\mathbf{w}, b, \boldsymbol{\xi}) = 1 - \xi_i - y_i(\mathbf{w}^T\mathbf{x}_i + b) \leq 0 \\ g_{2,i}(\mathbf{w}, b, \boldsymbol{\xi}) = -\xi_i \leq 0 \end{array} , \ i = 1, ..., N \\ \text{variables} & \mathbf{w} \in \mathbb{R}^m, b \in \mathbb{R}, \boldsymbol{\xi} \in \mathbb{R}^N \end{array} \tag{6}$$

as well its **Lagrangian dual problem**:

$$\begin{array}{ll} \text{maximize} & \theta(\alpha, \beta) = \inf_{\mathbf{w} \in \mathbb{R}^m, b \in \mathbb{R}, \boldsymbol{\xi} \in \mathbb{R}^N} \overbrace{\left\{ f(\mathbf{w}, b, \boldsymbol{\xi}) + \sum_{i=1}^{N} \alpha_i g_{1,i}(\mathbf{w}, b, \boldsymbol{\xi}) + \sum_{i=1}^{N} \beta_i g_{2,i}(\mathbf{w}, b, \boldsymbol{\xi}) \right\}}^{L(\mathbf{w}, b, \boldsymbol{\xi}, \alpha, \beta)} \\ \text{subject to} & \alpha_i \geq 0, \beta_i \geq 0, i = 1, ..., N \\ \text{variables} & \alpha_i \in \mathbb{R}, \beta_i \in \mathbb{R}, i = 1, ..., N \end{array} \tag{7}$$

where the Lagranian function can be explicitly written as

$$L(\mathbf{w}, b, \boldsymbol{\xi}, \alpha, \beta) = \frac{1}{2}\|\mathbf{w}\|^2 + \sum_{i=1}^{N} C_i \xi_i + \sum_{i=1}^{N} \alpha_i \left(1 - \xi_i - y_i(\mathbf{w}^T\mathbf{x}_i + b)\right) + \sum_{i=1}^{N} \beta_i \left(-\xi_i\right).$$

5.1.1 SVM Dual Problem in Explicit Form

Lemma 5.1. *Let (α, β) be a feasible solution to (7). If the following conditions hold:*

$$\sum_{i=1}^{N} \alpha_i y_i = 0, \ \ \alpha_i + \beta_i = C_i \ \forall i = 1, ..., N \tag{8}$$

then

$$\theta(\alpha, \beta) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j. \tag{9}$$

Otherwise, if (8) is not satisfied, then $\theta(\alpha, \beta) = -\infty$.

Proof. Take partial deriviates of the Lagranian function L over $\mathbf{w}, b, \boldsymbol{\xi}$, one has

$$\nabla_{\mathbf{w}} L = \mathbf{w} - \sum_{i=1}^{N} \alpha_i y_i \mathbf{x}_i, \quad \frac{\partial}{\partial b} L = -\sum_{i=1}^{N} \alpha_i y_i, \quad \frac{\partial}{\partial \xi_i} L = C_i - \alpha_i - \beta_i$$

If (8) is satisfied, then $\theta(\alpha,\beta) = L(\mathbf{w}, b, \boldsymbol{\xi}, \alpha, \beta)$ iff $\mathbf{w} = \sum_{i=1}^N \alpha_i y_i \mathbf{x}_i$, at which (9) holds; Otherwise, as $\frac{\partial}{\partial b}L$, $\frac{\partial}{\partial \xi_1}L$, ..., $\frac{\partial}{\partial \xi_N}L$ are constants not identically zero, one has $\theta(\alpha,\beta) = -\infty$. □

Remark 5.2. *The stationary condition $\theta(\alpha,\beta) = L(\mathbf{w}, b, \boldsymbol{\xi}, \alpha, \beta)$ holds iff (8) is satisfied and $\mathbf{w} = \sum_{i=1}^N \alpha_i y_i \mathbf{x}_i$.*

By Lemma.5.1, one may rewrite (7) as

$$\begin{array}{ll} \text{maximize} & \sum_{i=1}^N \alpha_i - \frac{1}{2}\sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \\ \text{subject to} & \alpha_i \geq 0, \beta_i \geq 0, \alpha_i + \beta_i = C_i, \sum_{i=1}^N \alpha_i y_i = 0, \quad i = 1, ..., N \\ \text{variables} & \alpha_i \in \mathbb{R}, \beta_i \in \mathbb{R}, i = 1, ..., N \end{array}$$

which can be further simplified as

$$\begin{array}{ll} \text{maximize} & \sum_{i=1}^N \alpha_i - \frac{1}{2}\sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \\ \text{subject to} & \sum_{i=1}^N \alpha_i y_i = 0 \\ \text{variables} & 0 \leq \alpha_i \leq C_i, \ i = 1, ..., N \end{array}$$

Claim 5.3. *There exists primal optimal solution $(\bar{\mathbf{w}},\bar{b},\bar{\boldsymbol{\xi}})$ to problem (6).*

Proof. (Warning: This proof requires basic understanding of mathematical analysis, which is beyond the scope of this course. It will not appear in course exams, and is merely a reference for those who are interested.)
WLOG one may assume the training data contains both positive and negative instances. [3] The set of all primal feasible solutions is (cf.(6))

$$\Omega=\bigcap_{i=1}^N \left(g_{1,i}^{-1}((-\infty,0])\cap g_{2,i}^{-1}((-\infty,0])\right),$$

which is an intersection of closed sets, so Ω is closed. Take some $(\mathbf{w}_0,b_0,\boldsymbol{\xi}_0)\in\Omega$, note that there exists compact set K where all primal feasible solutions $(\mathbf{w},b,\boldsymbol{\xi})\in\Omega$ with $f(\mathbf{w},b,\boldsymbol{\xi})\leq f(\mathbf{w}_0,b_0,\boldsymbol{\xi}_0)$ must satisfy $(\mathbf{w},b,\boldsymbol{\xi})\in K$.

- To see this, if $(\mathbf{w},b,\boldsymbol{\xi})\in\Omega$ and $f(\mathbf{w},b,\boldsymbol{\xi})\leq f(\mathbf{w}_0,b_0,\boldsymbol{\xi}_0)=\kappa_0$, then $\|\mathbf{w}\|\leq\sqrt{2\kappa_0}$, and that $|\xi_i|\leq\kappa_0/C_i$, $1-\xi_i-y_i(\mathbf{w}^T\mathbf{x}_i+b)\leq 0$ for all $i=1,...,N$. Note that for arbitrary positive instance (with index i) and negative instance (with index j), that is $y_i=-y_j=1$, one has

  $$1-\frac{\kappa_0}{C_i}-\sqrt{2\kappa_0}\|\mathbf{x}_i\|\leq 1-\xi_i-\mathbf{w}^T\mathbf{x}_i\leq b\leq\xi_j-1-\mathbf{w}^T\mathbf{x}_j\leq\frac{\kappa_0}{C_j}-1+\sqrt{2\kappa_0}\|\mathbf{x}_j\|$$

  which implies

  $$l=\max_{i:y_i=1}\left\{1-\frac{\kappa_0}{C_i}-\sqrt{2\kappa_0}\|\mathbf{x}_i\|\right\}\leq b\leq\min_{j:y_j=-1}\left\{\frac{\kappa_0}{C_j}-1+\sqrt{2\kappa_0}\|\mathbf{x}_j\|\right\}=u.$$

  Hence $(\mathbf{w},b,\boldsymbol{\xi})\in K$, for which $K=\left\{(\mathbf{w},b,\boldsymbol{\xi}):\|\mathbf{w}\|\leq\sqrt{2\kappa_0},|\xi_i|\leq\kappa_0/C_i,l\leq b\leq u\right\}$ is a compact set.

[3] Otherwise, say $y_1=\cdots=y_N=\zeta$ for some $\zeta\in\{\pm1\}$, then $(\mathbf{w},b,\boldsymbol{\xi})=(\mathbf{0},\zeta,\mathbf{0})$ is a trivial primal optimal solution.

Since f is continuous and $\Omega\cap K$ is compact, so $f(\Omega\cap K)$ is compact. Hence there exists $(\bar{\mathbf{w}},\bar{b},\bar{\boldsymbol{\xi}})\in\Omega\cap K$ for which

$$f(\bar{\mathbf{w}},\bar{b},\bar{\boldsymbol{\xi}})=\inf_{(\mathbf{w},b,\boldsymbol{\xi})\in\Omega\cap K} f(\mathbf{w},b,\boldsymbol{\xi})=\inf_{(\mathbf{w},b,\boldsymbol{\xi})\in\Omega} f(\mathbf{w},b,\boldsymbol{\xi})$$

In other words, $(\bar{\mathbf{w}},\bar{b},\bar{\boldsymbol{\xi}})$ is a primal optimal solution. ☐

As we have shown the existence of primal optimal solution $(\bar{\mathbf{w}},\bar{b},\bar{\boldsymbol{\xi}})$, Theorem.4.5 implies there also exists dual optimal solution $(\bar{\alpha},\bar{\beta})$. In the following claims we further link the relations between $(\bar{\mathbf{w}},\bar{b},\bar{\boldsymbol{\xi}})$ and $(\bar{\alpha},\bar{\beta})$.

5.1.2 KKT Condition for SVM

Recall Remark.5.2, the stationary condition holds iff

$$\begin{array}{ll} \text{(S1)} & \alpha_i+\beta_i=C_i \ \forall i=1,...,N \\ \text{(S2)} & \sum_{i=1}^N \alpha_i y_i = 0 \\ \text{(S3)} & \mathbf{w}=\sum_{i=1}^N \alpha_i y_i \mathbf{x}_i \end{array}$$

The primal and dual feasibility conditions are

$$\left.\begin{array}{ll} \text{(P1)} & g_{1,i}(\mathbf{w},b,\boldsymbol{\xi})=1-\xi_i-y_i(\mathbf{w}^T\mathbf{x}_i+b)\le 0 \\ \text{(P2)} & g_{2,i}(\mathbf{w},b,\boldsymbol{\xi})=-\xi_i\le 0 \\ \text{(D1)} & \alpha_i\ge 0 \\ \text{(D2)} & \beta_i\ge 0 \end{array}\right. \qquad \forall i=1,...,N$$

The complementary slackness condition is

$$\left.\begin{array}{ll} \text{(C1)} & \alpha_i g_{1,i}(\mathbf{w},b,\boldsymbol{\xi})=\alpha_i\left(1-\xi_i-y_i(\mathbf{w}^T\mathbf{x}_i+b)\right)=0 \\ \text{(C2)} & \beta_i g_{2,i}(\mathbf{w},b,\boldsymbol{\xi})=\beta_i(-\xi_i)=0 \end{array}\right. \quad \forall i=1,...,N$$

Claim 5.4. *Suppose $(\bar{\mathbf{w}},\bar{b},\bar{\boldsymbol{\xi}})$ and $(\bar{\alpha},\bar{\beta})$ are the optimal solutions to problems (6) and (7) respectively, then*

$$\bar{\mathbf{w}}=\sum_{i=1}^N \bar{\alpha}_i y_i \mathbf{x}_i, \quad \bar{b}=\arg\min_{b\in\mathbb{R}} \sum_{i=1}^N C_i \max\left(1-y_i(\bar{\mathbf{w}}^T\mathbf{x}_i+b),0\right).$$

$$\begin{cases} \bar{\alpha}_i=C_i, & \bar{\xi}_i=1-y_i(\bar{\mathbf{w}}^T\mathbf{x}_i+\bar{b}), & \text{if } y_i(\bar{\mathbf{w}}^T\mathbf{x}_i+\bar{b})<1 \\ \bar{\alpha}_i=0, & \bar{\xi}_i=0, & \text{if } y_i(\bar{\mathbf{w}}^T\mathbf{x}_i+\bar{b})>1 \\ 0\le\bar{\alpha}_i\le C_i, & \bar{\xi}_i=0, & \text{if } y_i(\bar{\mathbf{w}}^T\mathbf{x}_i+\bar{b})=1 \end{cases}$$

In particular, $\bar{\xi}_i=\max\left(1-y_i(\bar{\mathbf{w}}^T\mathbf{x}_i+\bar{b}),0\right)$ for all $i=1,...,N$.

Proof. Observe that (6) can be rewritten as the following optimization problem

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\|\mathbf{w}\|^2+\sum_{i=1}^N C_i\max\left(1-y_i(\mathbf{w}^T\mathbf{x}_i+b),0\right) \\ \text{variables} & \mathbf{w}\in\mathbb{R}^m, b\in\mathbb{R} \end{array}$$

Given the optimal weighting vector $\bar{\mathbf{w}}$, the optimal bias is given by

$$\bar{b} = \arg\min_{b \in \mathbb{R}} \sum_{i=1}^{N} C_i \max\left(1 - y_i(\bar{\mathbf{w}}^T \mathbf{x}_i + b), 0\right).$$

By Theorem.4.6, $(\bar{\mathbf{w}}, \bar{b}, \bar{\boldsymbol{\xi}}, \bar{\alpha}, \bar{\beta})$ satisfies KKT conditions, so $\bar{\mathbf{w}} = \sum_{i=1}^{N} \bar{\alpha}_i y_i \mathbf{x}_i$ and that the following satisfies for all $i = 1, ..., N$:

$$\begin{array}{ll}
\text{(P1)} & y_i(\bar{\mathbf{w}}^T \mathbf{x}_i + \bar{b}) \geq 1 - \bar{\xi}_i \\
\text{(P2)} & \bar{\xi}_i \geq 0 \\
\text{(D1)} & \bar{\alpha}_i \geq 0 \\
\text{(D2)} & \bar{\beta}_i \geq 0 \\
\text{(C1)} & \bar{\alpha}_i \left(1 - \bar{\xi}_i - y_i(\bar{\mathbf{w}}^T \mathbf{x}_i + \bar{b})\right) = 0 \\
\text{(C2)} & \bar{\beta}_i \bar{\xi}_i = 0 \\
\text{(S1)} & \bar{\alpha}_i + \bar{\beta}_i = C_i
\end{array}$$

- If $y_i(\bar{\mathbf{w}}^T \mathbf{x}_i + \bar{b}) < 1$, then $\bar{\xi}_i > 0$ by (P1), $\bar{\beta}_i = 0$ by (C2), $\bar{\alpha}_i = C_i$ by (S1), $\bar{\xi}_i = 1 - y_i(\bar{\mathbf{w}}^T \mathbf{x}_i + \bar{b})$ by (C1).
- If $y_i(\bar{\mathbf{w}}^T \mathbf{x}_i + \bar{b}) > 1$, then $\bar{\alpha}_i = 0$ by (P2,C1), so $\bar{\beta}_i = C_i$ by (S1), so $\bar{\xi}_i = 0$ by (C2).
- If $y_i(\bar{\mathbf{w}}^T \mathbf{x}_i + \bar{b}) = 1$, then $\bar{\xi}_i = 0$ by (S1,C1,C2), and $0 \leq \bar{\alpha}_i \leq C_i$ by (D1,D2,S1).

$\square$