1

Lagrangian multipliers:

$$\min_{\mathbf{w}} f(\mathbf{w})$$

$$h_i(\mathbf{w}) = 0 \qquad : \beta_i$$

$$L(\mathbf{w}, \beta_1, \ldots, \beta_n) = f(\mathbf{w}) + \beta_1 h_1(\mathbf{w}) + \ldots + \beta_n h_n(\mathbf{w})$$

Solution: $\left.\begin{array}{l} \dfrac{\partial L}{\partial \mathbf{w}} = 0 \\ \vdots \\ \dfrac{\partial L}{\partial \beta_i} = 0 \\ \vdots \end{array}\right\}$ system of $\underbrace{n+d}_{\text{the dimension of } \mathbf{w}}$ equations

Lagrangian duality, general case, also allowing inequalities:

$$\min_{\mathbf{w}} f(\mathbf{w})$$

$$g_i(\mathbf{w}) \le 0 \qquad : \alpha_i$$
$$h_j(\mathbf{w}) = 0 \qquad : \beta_j$$

$$L(\mathbf{w}, \alpha, \beta) = f(\mathbf{w}) + \sum_i \alpha_i g_i(\mathbf{w}) + \sum_j \beta_j h_j(\mathbf{w})$$

Primal problem: $\min_{\mathbf{w}} \theta_P(\mathbf{w})$ where $\theta_P(w) = \max_{\alpha, \beta : \alpha_i \ge 0} L(\mathbf{w}, \alpha, \beta)$

Dual problem: $\max_{\alpha, \beta : \alpha_i \ge 0} \theta_D(\alpha, \beta)$ where $\theta_D(\alpha, \beta) = \min_{\mathbf{w}} L(\mathbf{w}, \alpha, \beta)$

Weak duality: $P^* \ge D^*$ where $P^* = \min \max L$
$D^* = \max \min L$

②

Karush – Kuhn – Tucker (KKT) conditions: $\nearrow$ when is $P^* = D^*$?

$\Rightarrow$ if $\alpha_i^* > 0$, then $g_i(w^*) = 0$

#1 $\frac{\partial}{\partial w_i} L(w^*, \alpha^*, \beta^*) = 0, \forall i$

#2 $\frac{\partial}{\partial \beta_i} L(w^*, \alpha^*, \beta^*) = 0, \forall i$

#3 $\alpha_i^* g_i(w^*) = 0, \forall i$

#4 $g_i(w^*) \le 0, \forall i$

#5 $\alpha_i^* \ge 0, \forall i$

Special case of SVM, solve dual. 1) Find $\theta_D(\alpha, \beta) = \min_{\mathbf{w}} L(\mathbf{w}, \alpha, \beta)$:

$$\min_{\mathbf{w}, b} \frac{1}{2} \mathbf{w}^T \mathbf{w} \qquad // \text{ no } \beta_j$$

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) \ge 1 \Rightarrow 1 - y_i(\mathbf{w}^T \mathbf{x}_i + b) \le 0, \ \forall i \quad : \alpha_i$$

$$L(\mathbf{w}, \alpha, \beta) = f(\mathbf{w}) + \textstyle\sum_i \alpha_i g_i(\mathbf{w}) + \sum_j \beta_j h_j(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + \sum_{i=1}^N \alpha_i [1 - y_i(\mathbf{w}^T\mathbf{x}_i + b)]$$

Solution: $$\frac{\partial}{\partial \mathbf{w}} L(\mathbf{w}, b, \alpha) = 0 = \mathbf{w} + \sum_{i=1}^N \alpha_i [-y_i \mathbf{x}_i] \Rightarrow$$

$$\mathbf{w} = \sum_{i=1}^N \alpha_i y_i \mathbf{x}_i$$

$$\frac{\partial}{\partial b} L(\mathbf{w}, b, \alpha) = \sum_{i=1}^N \alpha_i y_i = 0$$

2) Plug solution for $\mathbf{w}$ into loss, find solution for α:

$$L(\mathbf{w}, b, \alpha) = \frac{1}{2}\Big[\sum_{j=1}^N \alpha_j y_j \mathbf{x}_j\Big]^T \Big[\sum_{i=1}^N \alpha_i y_i \mathbf{x}_i\Big]$$

$$+ \sum_{i=1}^N \alpha_i \Big[1 - y_i\Big(\Big(\sum_{j=1}^N \alpha_j y_j \mathbf{x}_j\Big)^T \mathbf{x}_i + b\Big)\Big]$$

$$= \frac{1}{2}\Big[\sum_{j=1}^N \alpha_j y_j \mathbf{x}_j\Big]^T\Big[\sum_{i=1}^N \alpha_i y_i \mathbf{x}_i\Big] + \sum_{i=1}^N \alpha_i - \Big[\sum_{j=1}^N \alpha_j y_j \mathbf{x}_j\Big]^T\Big[\sum_{i=1}^N \alpha_i y_i \mathbf{x}_i\Big] - \overbrace{\sum_{i=1}^N \alpha_i y_i b}^{0}$$

(3) $$= \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\Big[\sum_{j=1}^{N} \alpha_j y_j \mathbf{x}_j\Big]^T \Big[\sum_{i=1}^{N} \alpha_i y_i \mathbf{x}_i\Big]$$

Hence solve SVM dual as QP:

$$\max_{\boldsymbol{\alpha}} \quad \boldsymbol{\alpha}^T \mathbf{1} - \frac{1}{2} \boldsymbol{\alpha}^T \mathbf{H} \boldsymbol{\alpha} \qquad \text{where } H_{ij} = y_i y_j \mathbf{x}_i^T \mathbf{x}_j$$

$\vec{\mathbf{1}}$: a column vector of ones

$$\text{s.t.} \quad \sum_{i=1}^{N} \alpha_i y_i = 0$$

$$\alpha_i \geq 0, \ \forall i$$

QP is concave because $\mathbf{H}$ is positive semi-definite:

$\mathbf{H} = \mathbf{A}^T \mathbf{A}$ (where $\mathbf{A}$ is a matrix with columns $y_j \mathbf{x}_j$) so

$\boldsymbol{\alpha}^T \mathbf{H} \boldsymbol{\alpha} = \boldsymbol{\alpha}^T \mathbf{A}^T \mathbf{A} \boldsymbol{\alpha} = \|\mathbf{A}\boldsymbol{\alpha}\|^2 \geq 0, \ \forall \boldsymbol{\alpha} \Rightarrow \boldsymbol{\alpha}^T \mathbf{H} \boldsymbol{\alpha}$ convex

Why do we care about dual?

#1 It exposes structure: a) $\mathbf{w} = \sum_{i=1}^{N} \alpha_i y_i \mathbf{x}_i$

b) KKT#3 $\alpha_i g_i(\mathbf{w}) = 0 \Rightarrow$ if $\alpha_i > 0$ then $g_i(\mathbf{w}) = 0$

i.e. if $\alpha_i > 0$ then $1 - y_i(\mathbf{w}^T \mathbf{x}_i + b) = 0 \Rightarrow y_i(\mathbf{w}^T \mathbf{x}_i + b) = 1$

i.e. sample i is a SV and lies on the margin!

so $\mathbf{w} = \sum_{i \in SV} \alpha_i y_i \mathbf{x}_i$ and non-SV points don't matter, SV is the set of support vectors

#2 The kernel trick: Data only appears in pairs as $\mathbf{x}_i^T \mathbf{x}_j$:

$$\mathbf{w}^T \mathbf{x} + b = \sum_{i=1}^{N} \alpha_i y_i \underline{\mathbf{x}_i^T \mathbf{x}} + b \quad \text{(prediction for } \mathbf{x} \text{ before threshold)}$$