Dual Maximum Margin Optimization

Proposition: (The Maximum Margin Lagrangian Dual) Given the primal maximum margin optimization, [a] then the Lagrangian dual optimization for maximum margin classifiers is

$$\max_{\overline{\alpha}} \phi'(\overline{\alpha}) = \max_{\overline{\alpha}} \left(\sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j \overline{x}_i \bullet \overline{x}_j \right),$$

subject to the constraints

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

$$\alpha_i \geq 0,$$

with $i = 1, \ldots, l$.

[a] See lecture notes on maximum margin classifiers.

The Dual Decision Function

We also know that given the optimal Lagrangian multipliers $\overline{\alpha}^*$ we can construct both $\overline{w}^*$,

$$\overline{w}^* = \sum_{i=1}^{l} \alpha_i^* y_i \overline{x}_i,$$

and b^*,

$$b^* = \sum_{i=1}^{l} \alpha_i^* y_i \overline{x}_i \bullet \overline{x}_{sv+} - 1,$$

where we pick one support vector from the set of available support vectors,

$$(\overline{x}_{sv+}, +1) \in \{(\overline{x}_i, +1) \mid (\overline{x}_i, +1) \in D \text{ and } \alpha_i^* > 0\},$$

The identity for b^* follows directly from the KKT complimentarity condition with $\alpha_j^* > 0$ for some point $(\overline{x}_j, y_j) \in D$, then $y_j(\overline{w}^* \bullet \overline{x}_j - b^*) - 1 = 0$ or,

$$\begin{aligned} \overline{w}^* \bullet \overline{x}_j &= b^* + 1 \quad \text{if } y_j = +1, \\ \overline{w}^* \bullet \overline{x}_j &= b^* - 1 \quad \text{if } y_j = -1. \end{aligned}$$

Plugging in $\overline{w}^*$ and solving for b^* gives us our required result.

The Dual Decision Function

Putting this all together gives us,

$$\begin{aligned} \hat{f}(\overline{x}) &= \text{sign}(\overline{w}^* \bullet \overline{x} - b^*) \\ &= \text{sign}\left(\sum_{i=1}^{l} \alpha_i^* y_i \overline{x}_i \bullet \overline{x} - \sum_{i=1}^{l} \alpha_i^* y_i \overline{x}_i \bullet \overline{x}_{sv+} + 1\right). \end{aligned}$$

As we would expect, the dual decision function is completely determined by the Lagrangian multipliers $\overline{\alpha}^*$.

Observing that the decision function is completely determined by points $\overline{x}_i$ with $\alpha_i^* > 0$, we can say that the dual decision function is completely determined by the support vectors.

We consider the dual decision function a *support vector machine*.

Linear SVMs

Given

- a dot product space $\mathbb{R}^n$ as our data universe with points $\overline{x} \in \mathbb{R}^n$,
- some target function $f : \mathbb{R}^n \rightarrow \{+1, -1\}$,
- a labeled, linearly separable training set,

$$D = \{(\overline{x}_1, y_1), (\overline{x}_2, y_2), \ldots, (\overline{x}_l, y_l)\} \subseteq \mathbb{R}^n \times \{+1, -1\},$$

where $y_i = f(\overline{x}_i)$,

then compute a model $\hat{f} : \mathbb{R}^n \rightarrow \{+1, -1\}$ using D such that,

$$\hat{f}(\overline{x}) \cong f(\overline{x}),$$

for all $\overline{x} \in \mathbb{R}^n$.

Linear SVMs

Here we take as our models the linear support vector machines,

$$\hat{f}(\overline{x}) = \text{sign}\left(\sum_{i=1}^{l} \alpha_i^* y_i \overline{x}_i \bullet \overline{x} - \sum_{i=1}^{l} \alpha_i^* y_i \overline{x}_i \bullet \overline{x}_{sv+} + 1\right),$$

and we compute our support vector models with the Lagrangian dual optimization for maximum margin classifiers,

$$\overline{\alpha}^* = \underset{\overline{\alpha}}{\text{argmax}} \left(\sum_{i=1}^{l} \alpha_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j \overline{x}_i \bullet \overline{x}_j\right),$$

subject to the constraints

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

$$\alpha_i \geq 0,$$

where $i = 1, \ldots, l$.