Support Vector Machines

Study Notes March 19, 2011 Kenneth Chu

1 Support Vector Machines for Two-Class Classification

$$\mathcal{D} \;=\; \{\,(\mathbf{x}_1, y_1)\,,\,(\mathbf{x}_2, y_2)\,,\,\ldots\,,\,(\mathbf{x}_m, y_m)\,\}\,, \qquad \mathbf{x}_i \in \mathbb{R}^d, \qquad y_i \in \{\,-1,\,1\,\}$$

Reminder:

- Let $\mathbf{n}, \mathbf{z} \in \mathbb{R}^d$, with $\mathbf{n} \neq \mathbf{0}$, be given. The hyperplane $H_{\mathbf{n},\mathbf{z}} \subset \mathbb{R}^d$ with normal vector $\mathbf{n}$ and containing the point $\mathbf{z}$ is given by:

$$H_{\mathbf{n},\mathbf{z}} \;:=\; \{\;\mathbf{x} \in \mathbb{R}^d \;|\; \langle\,\mathbf{x} - \mathbf{z}\,,\,\mathbf{n}\,\rangle = 0\;\} \;=\; \left\{\;\mathbf{x} \in \mathbb{R}^d \;\middle|\; \left\langle\,\mathbf{x} - \mathbf{z}\,,\,\frac{\mathbf{n}}{\|\,\mathbf{n}\,\|}\,\right\rangle = 0\;\right\} \;=\; \{\;\mathbf{x} \in \mathbb{R}^d \;|\; \langle\,\mathbf{x} - \mathbf{z}\,,\,\widehat{\mathbf{n}}\,\rangle = 0\;\}$$

- Note that $H_{\mathbf{n},\mathbf{z}} = H_{\alpha\mathbf{n},\mathbf{z}}$, for any $\alpha \neq 0$.
- Let $\mathbf{x} \in \mathbb{R}^d$. The distance between $\mathbf{x}$ and the hyperplane $H_{\mathbf{n},\mathbf{z}}$ is given by:

$$\mathrm{dist}(\mathbf{x}, H_{\mathbf{n},\mathbf{z}}) \;=\; \left|\left\langle\,\mathbf{x} - \mathbf{z}\,,\,\frac{\mathbf{n}}{\|\mathbf{n}\|}\,\right\rangle\right|$$

  This implies:

$$\|\,\mathbf{n}\,\| \cdot \mathrm{dist}(\mathbf{x}, H_{\mathbf{n},\mathbf{z}}) \;=\; |\langle\,\mathbf{x} - \mathbf{z}\,,\,\mathbf{n}\,\rangle|$$

- Note that $\mathrm{dist}(\mathbf{x}, H_{\mathbf{n},\mathbf{z}})$ is well-defined, i.e. it depends only on the point $\mathbf{x}$ and the hyperplane $H_{\mathbf{n},\mathbf{z}}$, and is indeed independent of the particular choice of the normal vector $\mathbf{n}$ and the point $\mathbf{z} \in H_{\mathbf{n},\mathbf{z}}$.

Lemma 1.1 *Let $\mathbf{x}_1, \ldots, \mathbf{x}_m \in \mathbb{R}^d$ and a hyperplane $H_{\mathbf{n},\mathbf{z}} \subset \mathbb{R}^d$ be given, with $\mathbf{x}_i \notin H_{\mathbf{n},\mathbf{z}}$, for each $i = 1, \ldots, m$. Without loss of generality, we may assume (by rescaling $\mathbf{n}$, if necessary) that the normal vector $\mathbf{n}$ satisfies*

$$\min_{1 \leq i \leq m} |\,\langle\,\mathbf{x}_i - \mathbf{z}\,,\,\mathbf{n}\,\rangle\,| \;=\; 1$$

Proof Without loss of generality, we may assume (by rescaling $\mathbf{n}$, if necessary, while leaving $\mathbf{z}$ unchanged) that the normal vector $\mathbf{n}$ satisfies

$$\|\,\mathbf{n}\,\| \;=\; \frac{1}{\displaystyle\min_{1 \leq i \leq m} \{\,\mathrm{dist}(\,\mathbf{x}_i, H_{\mathbf{n},\mathbf{z}})\,\}}$$

With this choice of normal vector $\mathbf{n}$, we have

$$\min_{1 \leq i \leq m} |\,\langle\,\mathbf{x}_i - \mathbf{z}\,,\,\mathbf{n}\,\rangle\,| \;=\; \min_{1 \leq i \leq m} \{\,\|\,\mathbf{n}\,\| \cdot \mathrm{dist}(\mathbf{x}_i, H_{\mathbf{n},\mathbf{z}})\,\} \;=\; \|\,\mathbf{n}\,\| \cdot \min_{1 \leq i \leq m} \{\,\mathrm{dist}(\mathbf{x}_i, H_{\mathbf{n},\mathbf{z}})\,\} \;=\; 1$$

$\square$

Definition 1.2 *Let $\mathcal{D}_0 := \{\,\mathbf{x}_1, \ldots, \mathbf{x}_m\,\} \subset \mathbb{R}^d$, and $H \subset \mathbb{R}^d$ be a hyperplane, with $\mathbf{x}_1, \ldots, \mathbf{x}_m \notin H$. A representation $H_{\mathbf{n},\mathbf{z}} = H$ of H is said to be in $\mathcal{D}_0$-canonical form if*

$$\min_{1 \leq i \leq m} |\,\langle\,\mathbf{x}_i - \mathbf{z}\,,\,\mathbf{n}\,\rangle\,| \;=\; 1$$

Corollary 1.3 *A $\mathcal{D}$-separating hyperplane $H_{\mathbf{n},\mathbf{z}}$ in $\mathcal{D}_0$-canonical form satisfies:*

$$y_i \cdot \langle\,\mathbf{x}_i - \mathbf{z}\,,\,\mathbf{n}\,\rangle \;\geq\; 1, \qquad \text{for each } \; i = 1, \ldots, m.$$

References