Review of Sep 26

Duality of linear classifiers

Primal: w, b

Dual: α, b

Classification rule: $h(x) = \text{sign}(w \cdot x + b) = \text{sign}\Big(\underbrace{\sum_{i=1}^{m} \alpha_i y_i (x_i \cdot x)}_{= w \cdot x} + b\Big)$

Training Primal

$$P(w, b, \xi) = \min_{w, b, \xi \geq 0} \frac{1}{2} w \cdot w + C \sum_{i=1}^{m} \xi_i$$

$$\text{s.t. } y_1 (w \cdot x_i + b) \geq 1 - \xi_1$$

$$\vdots$$

$$y_m (w \cdot x_i + b) \geq 1 - \xi_m$$

Training Dual

$$D(\alpha) = \max_{\alpha \geq 0} \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j)$$

$$\text{s.t. } \sum_{i=1}^{m} \alpha_i = 0$$

Strong Duality: for any feasible w, b, ξ, α then $P(w, b, \xi) \geq D(\alpha)$
if $w^*, b^*, \xi^*, \alpha^*$ that are solution $P(w^*, b^*, \xi^*) = D(\alpha^*)$

Theorem: $err_{loo}(SVM) \leq \frac{1}{m} |\{i : \alpha_i > 0\}|$ = fraction of Support Vectors

Theorem: $err_{loo}(SVM) \leq \frac{1}{m} \frac{R^2}{\gamma^2}$ for homogeneous hard-margin SVM

Lemma (hard-margin homogeneous SVM): If (x_i, y_i) is a leave-one-out error, the $\alpha_i R^2 \geq 1$.

Proof: $err_{loo}(SVM) \leq \frac{1}{m} |\{i : \alpha_i \cdot R^2 \geq 1\}| \leq \frac{1}{m} \sum_{i=1}^{m} \alpha_i R^2 = \frac{1}{m} R^2 w \cdot w = \frac{1}{m} \frac{R^2}{\gamma^2}$ $\square$

Note: $P(w^*) = D(\alpha^*) \Leftrightarrow \frac{1}{2} w \cdot w = \sum \alpha_i - \underbrace{\frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j)}_{= \frac{1}{2} w \cdot w}$

$\Leftrightarrow w \cdot w = \sum \alpha_i$

Note: $\gamma = \frac{1}{\|w\|} = \frac{1}{\sqrt{w \cdot w}}$

Extending the feature space

- Mapping $\phi : X_I \to X_F$
- X_I called input space with dimensionality N_I (typically low)
- X_F called feature space with " N_F (typically high)

Example

Input space: $x = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$ → Feature space $\phi(x) = \begin{pmatrix} 4 \\ 9 \\ 2 \\ 3 \\ 6 \\ 1 \end{pmatrix}$

Separating w in feature space: $w = \begin{pmatrix} -0.7 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$

Example:

$$\phi\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right) = \begin{pmatrix} x_1^2 \\ x_2^2 \\ x_1\sqrt{2} \\ x_2\sqrt{2} \\ x_1 \cdot x_2\sqrt{2} \\ 1 \end{pmatrix}$$

$$\phi(\vec{a}) \cdot \phi(\vec{b}) = \phi\left(\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}\right) \cdot \phi\left(\begin{pmatrix} b_1 \\ b_2 \end{pmatrix}\right)$$

$$= \begin{pmatrix} a_1^2 \\ a_2^2 \\ a_1\sqrt{2} \\ a_2\sqrt{2} \\ a_1 \cdot a_2\sqrt{2} \\ 1 \end{pmatrix} \cdot \begin{pmatrix} b_1^2 \\ b_2^2 \\ b_1\sqrt{2} \\ b_2\sqrt{2} \\ b_1 \cdot b_2\sqrt{2} \\ 1 \end{pmatrix}$$

$$= a_1^2 \cdot b_1^2 + a_2^2 \cdot b_2^2 + 2a_1 \cdot b_1 + 2a_2 \cdot b_2 + 2a_1 a_2 b_1 b_2 + 1$$

$$= (a_1 \cdot b_1 + a_2 \cdot b_2 + 1)^2$$

$$= \left(\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \cdot \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} + 1\right)^2$$

$$= k(\vec{a}, \vec{b})$$

Working with kernels:

$$\|\phi(a) - \phi(b)\| = \sqrt{(\phi(a) - \phi(b))^2}$$

$$= \sqrt{\phi(a) \cdot \phi(a) - 2\phi(a)\phi(b) + \phi(b) \cdot \phi(b)}$$

$$= \sqrt{k(a,a) - 2k(a,b) + k(b,b)}$$

Example: $k(x,z) = \lambda\, k_1(x,z) + (1-\lambda)\, k_2(x,z)$

Symmetric: Yes, since Gram matrices K_1 and K_2 are symmetric

Pos semi-definite:

$$\forall \alpha_1 \dots \alpha_m : \sum_i \sum_j \alpha_i \alpha_j k_{ij} \geq 0$$

$$\Leftarrow \sum_i \sum_j \alpha_i \alpha_j k_{ij} = \sum_i \sum_j \alpha_i \alpha_j \left(\lambda k_1(x,z) + (1-\lambda) k_2(x,z) \right)$$

$$= \underbrace{\lambda \sum_i \sum_j \alpha_i \alpha_j k_1(x,z)}_{\geq 0} + (1-\lambda) \underbrace{\sum_i \sum_j \alpha_i \alpha_j k_2(x,z)}_{\geq 0}$$