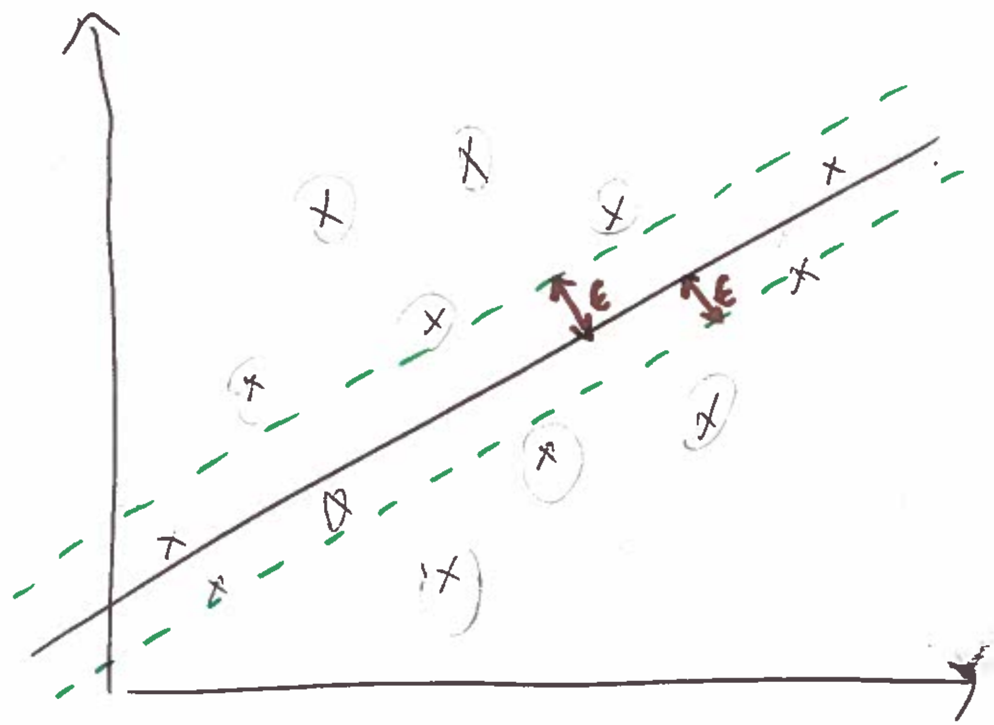
SVM for Regression

Epsilon Insensitive Loss Function



$$L_\epsilon(y,\hat{y}) \triangleq \begin{cases} 0 & \text{if } |y-\hat{y}|<\epsilon \\ |y-\hat{y}|-\epsilon & \text{otherwise} \end{cases}$$

Not differentiable because of the absolute value function

Using slack variables

$$y_i \leq f(x_i) + \epsilon + \xi_i^+$$

$$y_i \geq f(x_i) - \epsilon - \xi_i^-$$

$$J = C \cdot \sum_{i=1}^{N} (\xi_i^+ + \xi_i^-) + \frac{1}{2} \|w\|^2$$

$$\hat{w} = \sum_i \alpha_i x_i$$

$$\hat{y}(x) = \hat{w}_0 + \hat{w}^T x$$

$$\hat{y}(x) = \hat{w}_0 + \sum_i \alpha_i \underbrace{x_i^T x}_{\text{inner product}}$$

$$= \hat{w}_0 + \sum_i \alpha_i \kappa(x_i, x)$$

SVMs for classification

$y_i \in \{-1, 1\}$

$$f(x_i) = \log \frac{p(y_i = 1 \mid x_i, w)}{p(y_i = -1 \mid x_i, w)} = w^T x_i = \eta_i$$

$$p(y_i \mid x_i, w) = \text{sigm}(\underbrace{y_i}_{\text{actual output}} \underbrace{\eta_i}_{w^T x_i})$$

case 1: $y_i = +1 \quad \eta_i > 0 \rightarrow y_i \eta_i > 0 \rightarrow \text{sigm}(y_i \eta_i) > 0.5$

case 2: $y_i = +1 \quad \eta_i < 0 \rightarrow y_i \eta_i < 0 \rightarrow \text{sigm}(y_i \eta_i) < 0.5$

case 3: $y_i = -1 \quad \eta_i > 0 \rightarrow y_i \eta_i < 0 \rightarrow \text{sigm}(y_i \cdot \eta_i) < 0.5$

case 4: $y_i = -1 \quad \eta_i < 0 \rightarrow y_i \cdot \eta_i > 0 \rightarrow \text{sigm}(y_i \eta_i) > 0.5$

$$L_{nll} = -\log p(y|x,w) = -\log \text{sigm}(y \cdot m)$$

$$\left\{ \text{sigm}(a) = \frac{1}{1+e^{-a}} \right.$$

$$= -\log\left[\frac{1}{1+e^{-ym}}\right] = \log\left(1+e^{-ym}\right)$$

<u>Hinge Loss</u>

$$L_{hinge}(y, m) = \max(0, 1-ym)$$

$y = +1$ & $m > 0 \rightarrow ym > 0$

$y = +1$ & $m < 0 \rightarrow ym < 0$

$y = -1$ & $m > 0 \rightarrow ym < 0$

$y = -1$ & $m < 0 \rightarrow ym > 0$

(*) In the cases where $\left.\begin{array}{l} y=+1 \ \& \ m \geq 1 \text{ OR} \\ y=-1 \ \& \ m \leq -1 \end{array}\right\} ym \geq 1$

$$\max(0, 1-ym) = 0$$

(*) In the cases where $\left.\begin{array}{l} y=+1 \ \& \ 0<m<1 \text{ OR} \\ y=-1 \ \& \ -1<m<0 \end{array}\right\} 0<ym<1$

$$\max(0, 1-ym) = 1-ym$$

(*) In the cases where $\left.\begin{array}{l} y=+1 \ \& \ m<0 \text{ OR} \\ y=-1 \ \& \ m>0 \end{array}\right\} ym<0$

$$\max(0, 1-ym) = (1-ym) \text{ which is } > 1$$

Prediction

$$\hat{y}(x) = \text{sgn}(f(x)) = \text{sgn}(\hat{w}_0 + \hat{w}^T x)$$

$$\hat{y}(x) = \text{sgn}\Big(\hat{w}_0 + \sum_i \alpha_i \underbrace{x_i^T x}_{\text{inner product}}\Big)$$

$$\hat{y}(x) = \text{sgn}\Big(\hat{w}_0 + \sum_i \alpha_i K(x_i, x)\Big)$$