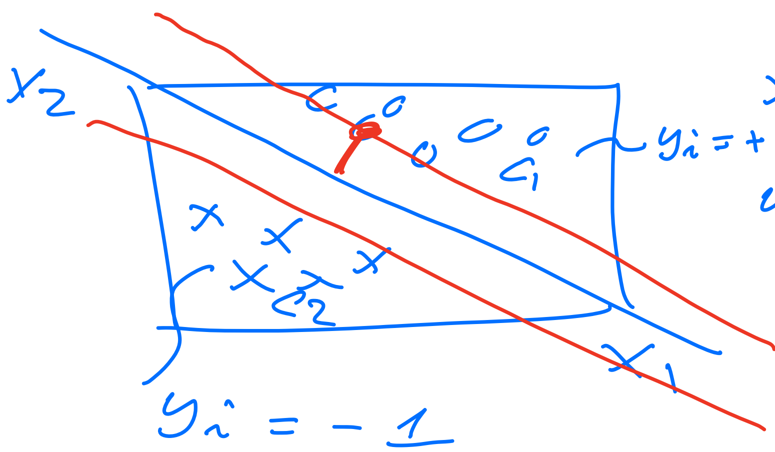
FYS-STK 4155, Nov 18, 2022

$$\mathcal{L}(w, w_0, \lambda) =$$

$$\frac{1}{2} w^T w - \sum_{i=0}^{n-1} \lambda_i \left(y_i (w^T x_i + w_0) - 1 \right)$$

$$\tilde{y}_i = w^T x_i + w_0$$



$$x^T = [x_1 \; x_2]$$

$$w^T = [w_1 \; w_2]$$

$$y_i (w^T x_i + w_0) = 1$$

Take derivatives

$$\frac{\partial \mathcal{L}}{\partial w} = 0 \Rightarrow w = \sum_{i=0}^{n-1} \lambda_i y_i x_i$$

$$\frac{\partial \mathcal{L}}{\partial w_0} = 0 \Rightarrow \sum_i \lambda_i y_i = 0$$

$$\vec{\mathcal{L}} = \sum_{i=0}^{n-1} \lambda_i - \frac{1}{2} \sum_{i,j} \lambda_i \lambda_j y_i y_j x_i^T x_j$$

with constraints (KKT)

$$\boxed{\begin{aligned} & y_i(w^T x_i + w_0) - 1 \geq 0 \\ & \lambda_i [y_i(w^T x_i + w_0) - 1] = 0 \\ & \lambda_i \geq 0 \end{aligned}}$$

$$\sum_i \lambda_i y_i = 0 \quad \leftarrow \frac{\partial \mathcal{L}}{\partial \lambda}$$

Rewrite Lagrangian as

$$\mathcal{L}(\lambda) = \mathbb{1}\lambda -$$

$$(\lambda^T = [\lambda_0, \lambda_1 \ldots \lambda_{n-1}]$$

$$- \frac{1}{2} \lambda^T P \lambda$$

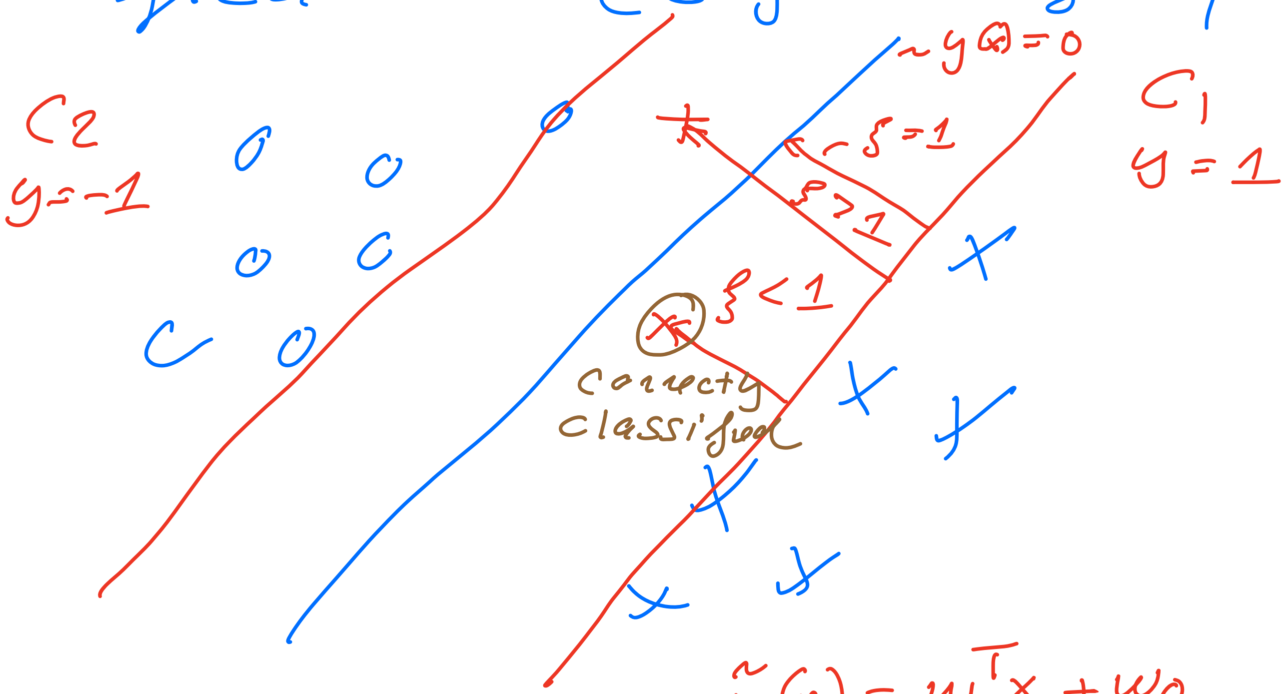
$$P = \begin{bmatrix} y_0 y_0 x_0^T x_0 & \cdots & y_0 y_{n-1} x_0^T x_{n-1} \\ \vdots & & \\ y_{n-1} y_0 x_{n-1}^T x_0 & \cdots & y_{n-1} y_{n-1} x_{n-1}^T x_{n-1} \end{bmatrix}$$

$$= \begin{bmatrix} y_0 y_0 k(x_0, x_0) & \cdots & y_0 y_{n-1} k(x_0, x_{n-1}) \\ \vdots & & \\ y_{n-1} y_0 k(x_{n-1}, x_0) & \cdots & \end{bmatrix}$$

$$K(x_i, x_j) = \phi^T(x_i)\,\phi(x_j)$$

$$\phi(x_i) = x_i \quad (\text{linear})$$

Margin with misclassification (soft margin)



$$\tilde{y}(x) = w^T x + w_0$$

$\tilde{y}(x) = 0$ on

the midpoint line

Define $\xi_i = 0$ for data points on the correct side of the margin

$\xi_i = |y_i - \tilde{y}_i|$ for other points

$\xi_i > 1$, misclassified

$\xi_i \leq 1$, classified correctly

We can replace (hard margin)

$$y_i(w^T x_i + w_0) \geq 1$$

$$y_i \tilde{y}_i = y_i f(x_i) \geq 1$$

replaced with

$$y_i(w^T x_i + w_0) \geq 1 - \xi_i$$

$$\forall i \quad i = 0, 1, 2, \ldots n-1$$

need to satisfy $\xi_i \geq 0$

$0 < \xi_i \leq 1$ points which are inside the margin but on the correct side,

$\xi_i > 1$ are misclassified points

we want to minimize

$$C \sum_{i=0}^{n-1} \xi_i + \frac{1}{2} w^T w$$

$C > 0$, controls trade-off between the slack parameters ξ_i and the margin

$$\mathcal{L}(w, w_0, \xi, \lambda, \mu) =$$

$$\frac{1}{2} w^T w + C \sum_i \xi_i -$$

↑ hyperparameter

$$- \sum_{i=0}^{n-1} \lambda_i \left(y_i (w^T x_i + w_0) - 1 + \xi_i \right)$$

$$- \sum_i \mu_i \xi_i$$

Lagrange multipliers

$$\lambda_i \geqslant 0 \quad \wedge \quad \mu_i \geqslant 0$$

The KKT conditions:

$$\lambda_i \geqslant 0$$

$$y_i (w^T x_i + w_0) - 1 + \xi_i \geqslant 0$$

$$\lambda_i \left[y_i (w^T x_i + w_0) - 1 + \xi_i \right] = 0$$

$$\mu_i \geqslant 0$$

$$\xi_i \geqslant 0$$

$$\mu_i \xi_i = 0$$

Rewrite $\mathcal{L}(w, w_0, \xi, \lambda, \mu)$ in just $\mathcal{L}(\lambda)$

$$\frac{\partial \mathcal{L}}{\partial w} = 0 \Rightarrow w = \sum_i \lambda_i y_i \underbrace{\Phi(x_i)}_{x_i}$$

$$\left(\tilde{y}_i = w^T x_i + w_0 \right.$$
$$\left. = w^T \phi(x_i) + w_0 \right)$$

$$\frac{\partial \mathcal{L}}{\partial w_0} = 0 \Rightarrow \sum_i \lambda_i y_i = 0$$

$$\frac{\partial \mathcal{L}}{\partial \xi_i} = 0 \Rightarrow \lambda_i = C - \mu_i$$

$\Rightarrow$

$$\mathcal{L}(\lambda) = \sum_{i=0}^{m-1} \lambda_i -$$

$$\frac{1}{2} \sum_{i,j} \lambda_i \lambda_j y_i y_j \underbrace{x_i^T x_j}_{K(x_i, x_j)}$$

<u>Note</u>

$\lambda_i \geq 0$ and $\mu_i \geq 0 \Rightarrow$

$$\lambda_i \leq C$$

we have to minimize wrt a constraints (with fixed C)

$$0 \leq \lambda_i \leq C$$

$$\sum_{i=0}^{n-1} \lambda_i y_i = 0$$

Need to satisfy

$$y_i(w^T x_i + w_0) = 1 - \xi_i$$

$$\xi_i \mu_i = 0$$

if $\mu_i > 0$ then $\xi_i = 0$

(we are on the margin)

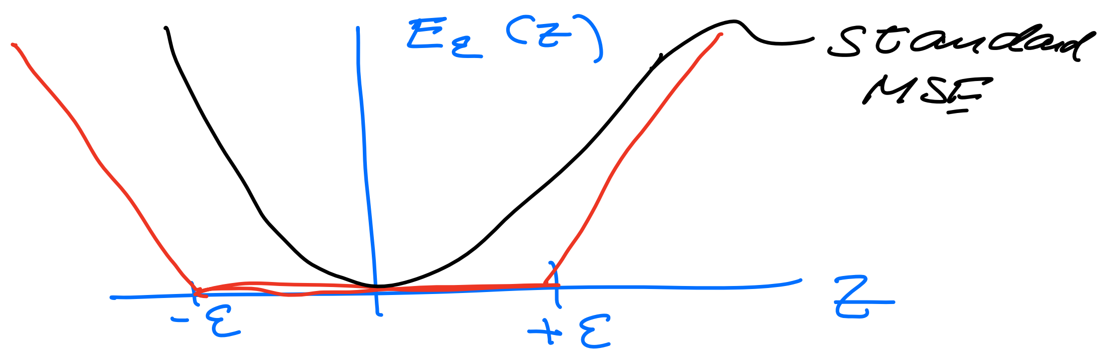
if $\mu_i = 0$, then $\xi_i \geq 0$

SVM for Regression

$$C = \frac{1}{n} \sum_{i=0}^{n-1} (y_i - \tilde{y}_i)^2 + \lambda w^T w$$

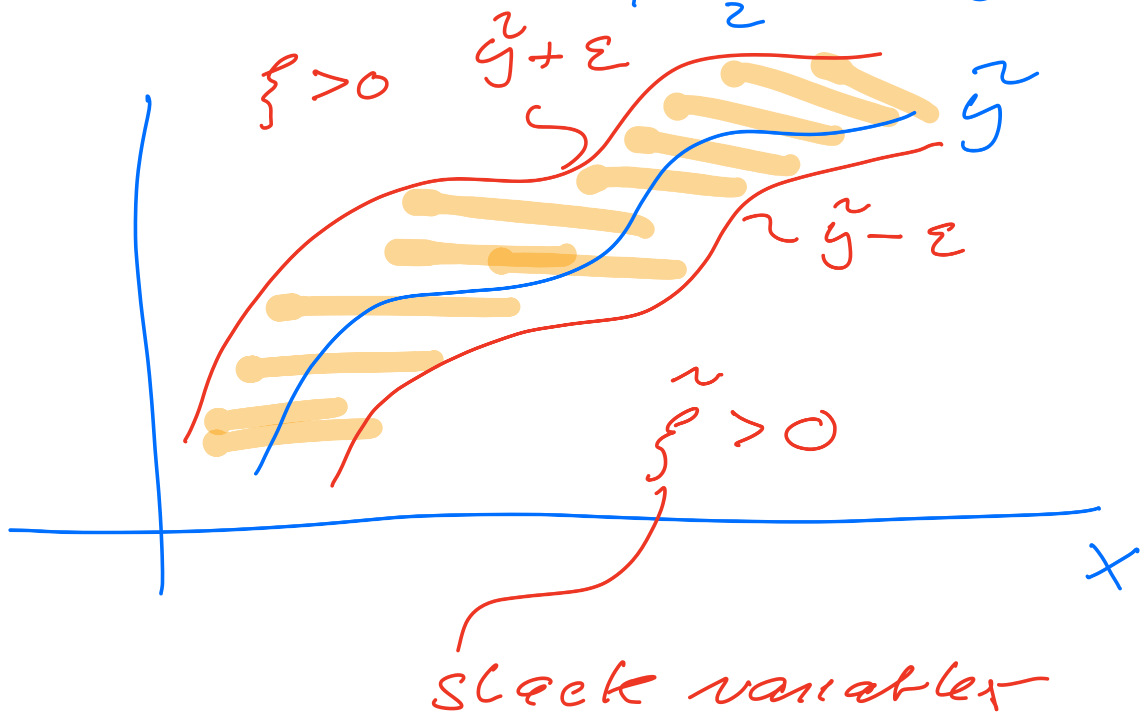
(Ridge regression)
SVMs deal with sparse
(few points defining)
solutions to the above
function;
Replace -C- by a ε-insensitive cost function
which gives zero if the
absolute value of the
difference between $\hat{y}$
and y is less than ε
$(\varepsilon > 0)$

$$E_\varepsilon(y-\hat{y}) = \begin{cases} 0 & \text{if } |y-\hat{y}| < \varepsilon \\ |y-\hat{y}| - \varepsilon & \text{otherwise} \end{cases}$$



define again $\tilde{y}_i = w^T\phi(x_i)$
$+ w_0$

$$C = \beta \sum_{i=0}^{n-1} E_\varepsilon(y_i - \tilde{y}_i) + \frac{1}{2} w^T w$$



For each x_i we have two

variables $\xi_i \geq 0$ and $\tilde{\xi}_i \geq 0$

when $y_i > \tilde{y}_i + \varepsilon$ $(\xi_i > 0)$

$\tilde{\xi}_i > 0$ corresponds to

$$y_i < \tilde{y}_i - \varepsilon$$

Need conditions for a target point y_i to be inside the ε-tube