Lecture 11: Support Vector Regression, Dual, Optimization Algorithm and Kernel Trick

Instructor: Prof. Ganesh Ramakrishnan

Formulation for Support Vector Regression

- $\min_{w,b,\xi_i,\xi_i^*} \frac{1}{2}\|w\|^2 + C\sum_i(\xi_i + \xi_i^*)$
  s.t. $\forall i$,
  $y_i - w^\top\phi(x_i) - b \leq \epsilon + \xi_i$,
  $b + w^\top\phi(x_i) - y_i \leq \epsilon + \xi_i^*$,
  $\xi_i, \xi_i^* \geq 0$
- Let's consider the lagrange multipliers α_i, α_i^*, μ_i and μ_i^* corresponding to the above-mentioned constraints respectively.

Consider Support Vector Regression Applet at
https://www.csie.ntu.edu.tw/~cjlin/libsvm/

KKT conditions

Recall: Regression curve: $f(x) = w^T \phi(x) + b$

- Differentiating the Lagrangian w.r.t. w,
  $w - \alpha_i \phi(x_i) + \alpha_i^* \phi(x_i) = 0$
  i.e. $w = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) \phi(x_i)$ $\rightarrow$ $w = \sum_{\alpha_i^* = 0} \alpha_i \phi(x_i) - \sum_{\alpha_i = 0} \alpha_i^* \phi(x_i)$
- Differentiating the Lagrangian w.r.t. ξ_i,
  $C - \alpha_i - \mu_i = 0$
  i.e. $\underline{\alpha_i + \mu_i = C}$ $\qquad \mu_i = C - \alpha_i$
- Differentiating the Lagrangian w.r.t ξ_i^*,
  $\underline{\alpha_i^* + \mu_i^* = C}$ $\qquad \mu_i^* = C - \alpha_i^*$
- Differentiating the Lagrangian w.r.t b,
  $\sum_i (\alpha_i^* - \alpha_i) = 0$
- Complimentary slackness:
  $\alpha_i (y_i - w^\top \phi(x_i) - b - \epsilon - \xi_i) = 0$
  $\underline{\mu_i \xi_i = 0}$ $\qquad (C - \alpha_i) \xi_i = 0$ $\qquad \xi_i > 0 \Rightarrow \alpha_i = C$
  $\alpha_i^* (b + w^\top \phi(x_i) - y_i - \epsilon - \xi_i^*) = 0$
  $\underline{\mu_i^* \xi_i^* = 0}$ $\qquad (C - \alpha_i^*) \xi_i^* = 0$ $\qquad \xi_i^* > 0 \Rightarrow \alpha_i^* = C$

Conclusions from the KKT conditions:

If $\xi_i > 0 \Rightarrow \xi_i^* = 0$
If $\xi_i^* > 0 \Rightarrow \xi_i = 0$

$y_i - w^T\phi(x_i) - b = \epsilon + \xi_i$ when x_i is "above" band and not o/w

$$\alpha_i(y_i - w^\top \phi(x_i) - b - \epsilon - \xi_i) = 0$$

and

$b + w^T\phi(x_i) - y_i = \epsilon + \xi_i^*$ when x_i is "below" band and not o/w

$$\alpha_i^*(b + w^\top \phi(x_i) - y_i - \epsilon - \xi_i^*) = 0$$

$\Rightarrow$? If $\alpha_i > 0 \Rightarrow y_i - w^T\phi(x_i) - b = \epsilon + \xi_i$
$\Rightarrow b + w^T\phi(x_i) - y_i < \epsilon + \xi_i^*$
$\Rightarrow \alpha_i^* = 0$
Similarly, if $\alpha_i^* > 0 \Rightarrow \alpha_i = 0$
So: in summary $\alpha_i \alpha_i^* = 0$ (A point lies either above or below the line)

Conclusions from the KKT conditions:

$$\alpha_i(y_i - w^T\phi(x_i) - b - \epsilon - \xi_i) = 0$$

$$\Downarrow$$

$$y_i - w^T\phi(x_i) - b = \epsilon + \cancel{\xi_i}^{\,0} = \epsilon$$

$\alpha_i \in (0, C) \Rightarrow ?$ $C - \alpha_i \in (0, C)$

$\underbrace{(C - \alpha_i)}_{\mu_i}\xi_i = 0 \Rightarrow ?$ $\xi_i = 0$

i.e: $\alpha_i \in (0,C)$ for points **ON** ϵ band

$\alpha_i^* \in (0, C) \Rightarrow ?$ $C - \alpha_i^* \in (0, C)$

$\underbrace{(C - \alpha_i^*)}_{\mu_i^*}\xi_i^* = 0 \Rightarrow ?$ $\xi_i^* = 0$

Similarly:

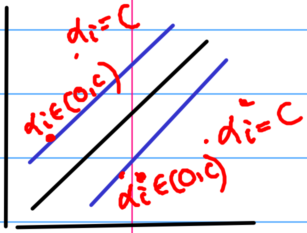
$\alpha_i^* \in (0,C)$ for points **ON** ϵ band

By elimination: $\alpha_i = \alpha_i^* = 0$ for points lying **WITHIN** the ϵ-band

In Summary: (1) $\alpha_i = C$ or $\alpha_i^* = C$ for points outside ϵ-band

(2) $\alpha_i \in (0, C)$ for points on ϵ band

(3) $\alpha_i = \alpha_i^* = 0$ for points within ϵ band



Points with $\alpha_i \in (0, C)$ are called support vectors

ONE Reason: Addition or deletion of points with $\alpha_i = \alpha_i^* = 0$ does not change the regression line

But points with $\alpha_i \in (0, C)$ or $\alpha_i^* \in (0, C)$ define the ϵ-band

Recall a KKT condition

$$w = \sum_i (\alpha_i - \alpha_i^*) \phi(x_i)$$

$$f(x) = w^T \phi(x) + b$$

$$= \sum_i (\alpha_i - \alpha_i^*) \phi^T(x_i) \phi(x) + b$$

For computing b: Recap that for all

$\alpha_i \in (0, C)$, $\quad y_i - w^T \phi(x_i) - b = \epsilon$

$$\Rightarrow b = \epsilon - y_i + w^T \phi(x_i) \text{ for any } \alpha_i \in (0, C)$$

$$\Rightarrow f(x) = \sum_i (\alpha_i - \alpha_i^*) \phi^T(x_i) \phi(x) + \underbrace{\left(\epsilon - y_j + w^T \phi(x_j)\right)}_{\text{for any } j \text{ s.t. } \alpha_j \in (0, C)}$$

$$\left(\epsilon - y_j + \sum_i (\alpha_i - \alpha_i^*) \phi^T(x_i) \phi(x_j)\right) = \text{for any } j \text{ s.t. } \alpha_j \in (0, C)$$

Derivation of dual:

$$\min_{w,b,\xi_i,\xi_i^*} \frac{1}{2}\|w\|^2 + C\sum_i (\xi_i + \xi_i^*)$$
$$+ \sum \alpha_i (y_i - \phi^T(x_i)w - b - \epsilon - \xi_i)$$
$$+ \sum \alpha_i^* (\phi^T(x_i)w + b - y_i - \epsilon - \xi_i^*)$$
$$- \sum_i \mu_i \xi_i - \sum_i \mu_i^* \xi_i^*$$

Substituting from KKT conditions

$$w = \sum (\alpha_i - \alpha_i^*)\phi(x_i), \quad \alpha_i + \mu_i = C, \quad \sum_i (\alpha_i - \alpha_i^*) = 0$$

$$\min_{w,\xi_i,\xi_i^*,b} \frac{1}{2}\|w\|^2 + \underbrace{b\sum_i (\alpha_i^* - \alpha_i)}_{=0} + \sum_i \xi_i \underbrace{(C - \alpha_i - \mu_i)}_{=0}$$
$$+ \sum_i \xi_i^* \underbrace{(C - \alpha_i^* - \mu_i^*)}_{=0}$$

$$- \frac{1}{2}\sum_i (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)\phi^T(x_i)\phi(x_j) + \sum y_i(\alpha_i - \alpha_i^*).$$

- The primal objective and constraints are convex $\Rightarrow$ KKT conditions here necessary and sufficient and strong duality holds
- $w = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)\phi(x_i)$ $\Rightarrow$ the final decision function $f(x) = w^T\phi(x) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)\phi^T(x_i)\phi(x)$ $+ (\epsilon - y_j + \sum_i (\alpha_i - \alpha_i^*)\phi^T(x_i)\phi(x_j))$
  $+b$
- The dual optimization problem to compute the α's for SVR is:

$$max_{\alpha_i, \alpha_i^*} - \frac{1}{2}\sum_i \sum_j (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)\phi^\top(x_i)\phi(x_j)$$

for some j s.t $\alpha_j \in (0, C)$

$$-\epsilon \sum_i (\alpha_i + \alpha_i^*) + \sum_i y_i(\alpha_i - \alpha_i^*)$$

s.t.

  - $\sum_i(\alpha_i - \alpha_i^*) = 0$
  - $\alpha_i, \alpha_i^* \in [0, C]$

  KKT constraints

- ***We notice that the only way these three expressions involve ϕ is through $\phi^\top(x_i)\phi(x_j) = K(x_i, x_j)$, for some i, j***