Lecture 11: Support Vector Regression, Dual 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.
- Aside: Consider Support Vector Regression Applet at `https://www.csie.ntu.edu.tw/~cjlin/libsvm/`.

KKT conditions

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

Conclusions from the KKT conditions:

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

and

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

$\Rightarrow$?

Conclusions from the KKT conditions:

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

$$(C - \alpha_i)\xi_i = 0 \Rightarrow ?$$

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

$$(C - \alpha_i^*)\xi_i^* = 0 \Rightarrow ?$$

For Support Vector Regression, since the original objective and the constraints are convex, any $(\mathbf{w}, \mathbf{b}, \alpha, \alpha^*, \mu, \mu^*, \xi, \xi^*)$ that satisfy the necessary KKT conditions gives optimality (conditions are also sufficient)

Some observations

- $\alpha_i, \alpha_i^* \geq 0$, $\mu_i, \mu_i^* \geq 0$, $\alpha_i + \mu_i = C$ and $\alpha_i^* + \mu_i^* = C$
  Thus, $\alpha_i, \mu_i, \alpha_i^*, \mu_i^* \in [0, C]$, $\forall i$
- If $0 < \alpha_i < C$, then $0 < \mu_i < C$
  (as $\alpha_i + \mu_i = C$)
- $\mu_i \xi_i = 0$ and $\alpha_i(y_i - -w^\top \phi(x_i) - b - \epsilon - \xi_i) = 0$ are complementary slackness conditions
  So $0 < \alpha_i < C \Rightarrow \xi_i = 0$ and $y_i - w^\top \phi(x_i) - b = \epsilon + \xi_i = \epsilon$
  - All such points lie on the boundary of the ϵ band
  - Using any point x_j (that is with $\alpha_j \in (0, C)$) on margin, we can recover b as:
    $b = y_j - w^\top \phi(x_j) - \epsilon$

Support Vector Regression
Dual Objective

Dual function

- Let $L^*(\alpha, \alpha^*, \mu, \mu^*) = \min_{w,b,\xi,\xi^*} L(w, b, \xi, \xi^*, \alpha, \alpha^*, \mu, \mu^*)$
- By weak duality theorem, we have:
  $\min_{w,b,\xi,\xi^*} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \geq L^*(\alpha, \alpha^*, \mu, \mu^*)$
  s.t. $y_i - w^\top \phi(x_i) - b \leq \epsilon - \xi_i$, and
  $w^\top \phi(x_i) + b - y_i \leq \epsilon - \xi_i^*$, and
  $\xi_i, \xi^* \geq 0$, $\forall i = 1, \ldots, n$
- The above is true for any $\alpha_i, \alpha_i^* \geq 0$ and $\mu_i, \mu_i^* \geq 0$
- Thus,

$$\min_{w,b,\xi,\xi^*} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \geq \max_{\alpha,\alpha^*,\mu,\mu^*} L^*(\alpha, \alpha^*, \mu, \mu^*)$$

s.t. $y_i - w^\top \phi(x_i) - b \leq \epsilon - \xi_i$, and
$w^\top \phi(x_i) + b - y_i \leq \epsilon - \xi_i^*$, and
$\xi_i, \xi^* \geq 0$, $\forall i = 1, \ldots, n$

Dual objective

- In case of Support Vector Regression, we have a strictly convex objective and linear constraints $\Rightarrow$ KKT conditions are necessary and sufficient and strong duality holds:

$$\min_{w,b,\xi,\xi^*} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) = \max_{\alpha,\alpha^*,\mu,\mu^*} L^*(\alpha, \alpha^*, \mu, \mu^*)$$

s.t. $y_i - w^\top\phi(x_i) - b \leq \epsilon - \xi_i$, and
$w^\top\phi(x_i) + b - y_i \leq \epsilon - \xi_i^*$, and
$\xi_i, \xi^* \geq 0$, $\forall i = 1, \ldots, n$

- This value is precisely obtained at the $(\mathbf{w}, \mathbf{b}, \xi, \xi^*, \alpha, \alpha^*, \mu, \mu^*)$ that satisfies the necessary (and sufficient) optimality conditions
- Given strong duality, we can equivalently solve

$$\max_{\alpha,\alpha^*,\mu,\mu^*} L^*(\alpha, \alpha^*, \mu, \mu^*)$$

- $L(\alpha, \alpha^*, \mu, \mu^*) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) +$
  $\sum_{i=1}^{n}\left(\alpha_i(y_i - w^\top\phi(x_i) - b - \epsilon - \xi_i) + \alpha_i^*(w^\top\phi(x_i) + b - y_i - \epsilon - \xi_i^*)\right)$
  $\sum_{i=1}^{n}(\mu_i\xi_i + \mu_i^*\xi_i^*)$
- We obtain w, b, ξ_i, ξ_i^* in terms of α, α^*, μ and μ^* by using the KKT conditions derived earlier as $w = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)\phi(x_i)$ and $\sum_{i=1}^{n}(\alpha_i - \alpha_i^*) = 0$ and $\alpha_i + \mu_i = C$ and $\alpha_i^* + \mu_i^* = C$
- Thus, we get:
  $L(w, b, \xi, \xi^*, \alpha, \alpha^*, \mu, \mu^*)$
  $= \frac{1}{2}\sum_i\sum_j(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)\phi^\top(x_i)\phi(x_j) +$
  $\sum_i\left(\xi_i(C - \alpha_i - \mu_i) + \xi_i^*(C - \alpha_i^* - \mu_i^*)\right) - b\sum_i(\alpha_i - \alpha_i^*) -$
  $\epsilon\sum_i(\alpha_i + \alpha_i^*) + \sum_i y_i(\alpha_i - \alpha_i^*) - \sum_i\sum_j(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)\phi^\top(x_i)\phi(x_j)$
  $= -\frac{1}{2}\sum_i\sum_j(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)\phi^\top(x_i)\phi(x_j) - \epsilon\sum_i(\alpha_i + \alpha_i^*) +$
  $\sum_i y_i(\alpha_i - \alpha_i^*)$

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

- $w = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)\phi(x_i) \Rightarrow$ the final decision function $f(x) = w^T\phi(x) + b =$ $\sum_{i=1}^{n}(\alpha_i - \alpha_i^*)\phi^T(x_i)\phi(x) + y_j - \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)\phi^T(x_i)\phi(x_j) - \epsilon$
  x_j is any point with $\alpha_j \in (0, C)$
- The dual optimization problem to compute the α's for SVR is:

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

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

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