Support Vector machine

Nonlinear case

Professor Abdellatif El Afia

Plan

- $\boldsymbol{C-SVC}$ for Nonlinear separation
- $\boldsymbol{\varepsilon}$-band SVR for Nonlinear Regression

$\boldsymbol{C-SVC}$ for nonlinear separation(NS)

Polynomial transformation

- $\phi_Q : X \subseteq \mathbb{R}^d \rightarrow Z(Hilbert\ space)$ such that $\phi_Q(x) = \begin{pmatrix} 1, \\ x_1, \dots, x_d, \\ x_1^2, x_1x_2, \dots, x_d^2, \\ \dots \\ x_1^Q, x_1^{Q-1}x_2, \dots, x_d^Q \end{pmatrix}$

The optimization Primal problem becomes

- $(\phi_Q(x_i), y_i)$

- $C - SVCNS \begin{cases} Min & \frac{1}{2}\|w\|^2 + C \sum_{i=1}^{n} \xi_i \\ s.t & y_i(w^T \boldsymbol{\phi_Q(x_i)} + b) \geq 1 - \xi_i, i = 1,..,n \\ & \xi_i \geq 0, w \in \mathbb{R}^d, b \in \mathbb{R} \end{cases}$

$C-SVC\mathbf{NS}$: Algorithm

- Input: training set : $\{(x_i, y_i\}_{i=1}^{n}$where $x_i \in \mathbb{R}^d, y_i \in \{1, -1\}$
- Choose an **approriate** $\boldsymbol{Q}$ and penalty parameter $C > 0$
- Contruct and solve the optimzation problem $C-SVC\mathbf{NS}$ obtaining (w^*, b^*, ξ^*), $\xi^* = (\xi_1^*, \dots, \xi_n^*)$

$$Primal: C-SVC\mathbf{NS} \begin{cases} Min & \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i \\ s.t & y_i(w^T\phi_Q(x_i) + b) \geq 1 - \xi_i, i = 1, .., n \\ & \xi_i \geq 0, w \in \mathbb{R}^d, b \in \mathbb{R} \end{cases}$$

- Contruit the separating hyperplane $(w^*)^T\phi_Q(x) + b^*$ and the decision function is

$$h_S\left(\phi_Q(x)\right) = sign((w^*)^T\phi_Q(x) + b^*) \rightarrow L_S(h_S) = \frac{1}{n}\sum_{i=1}^{n}1_{\left\{h_{w^*,b^*}\left(\phi_Q(x_i)\right)\neq y_i\right\}}$$

$C - SVC\mathbf{NS}$: Dual form

Dual form :

$$Dual: C - SVCNS \begin{cases} Max \quad \sum_{i=1}^{n} \lambda_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_j \lambda_i y_j y_i \left(\phi_Q(x_j)^T \phi_Q(x_i) \right) \\ s.t \quad \sum_{i=1}^{n} \lambda_i y_i = 0 \\ \quad C - \lambda_i - \mu_i = 0, i = 1, \dots, n \\ \quad \mu_i \geq 0, \lambda_i \geq 0 \; i = 1, \dots, n \end{cases}$$

$C - SVC\mathbf{NS}$: Dual form

$$
Dual: C - SVCNS \Leftrightarrow \begin{cases} Max \quad \sum_{i=1}^{n} \lambda_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_j \lambda_i y_j y_i \left(\boldsymbol{\phi}_{\boldsymbol{Q}}(\boldsymbol{x}_j)^{\boldsymbol{T}} \boldsymbol{\phi}_{\boldsymbol{Q}}(\boldsymbol{x}_i)\right) \\ s.t \qquad \sum_{i=1}^{n} \lambda_i y_i = 0 \\ \qquad C \geq \lambda_i \geq 0 \; i = 1, \dots, n \end{cases}
$$

$C - SVC$**NS** : Dual-Algorithm

- Input: training set : $\{(x_i, y_i\}_{i=1}^{n}$ where $x_i \in \mathbb{R}^d, y_i \in \{1, -1\}$
- Choose an **approriate map ϕ_Q** and penalty parameter $C > 0$
- Contruct and solve the optimzation problem Dual: $\mathrm{C} - SVC$ obtaining $\lambda^* = (\lambda_1^*, \dots, \lambda_n^*)$

$$Dual: C - SVCNS \begin{cases} Max \quad \sum_{i=1}^{n} \lambda_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_j \lambda_i y_j y_i \left(\phi_Q(x_j)^T \phi_Q(x_i) \right) \\ s.t \quad \sum_{i=1}^{n} \lambda_i y_i = 0 \\ C \geq \lambda_i \geq 0 \,, i = 1, \dots, n \end{cases}$$

- Choose a possitive component of λ^*, $\lambda_j^* \in (0, C)$, and Compute

  $w^* = \sum_{i=1}^{n} \lambda_i^* \, y_i \phi_Q(x_i)$ and $b^* = y_j - \sum_{i=1}^{n} \lambda_i^* y_i \left(\phi_Q(x_j)^T \phi_Q(x_i) \right)$
- Contruit the separating hyperplane $(w^*)^T x + b^*$ and the decision function is

$$h(x) = sign((w^*)^T x + b^*)$$

ε-band SVR for Nonlinear Regression(NR)

Polynomial transformation

- $\phi_Q : X \to Z$ such that $\phi_Q(x) = \begin{pmatrix} 1, \\ x_1, \ldots, x_d, \\ x_1^2, x_1 x_2, \ldots, x_d^2, \\ \ldots \\ x_1^Q, x_1^{Q-1} x_2, \ldots, x_d^Q \end{pmatrix}$ Primal Problem

$$\text{PLS} - \varepsilon\text{–band} - SVR\text{NR}: \begin{cases} \min \quad \frac{1}{2} w^T w + C \sum_{i=1}^{n} (\xi_i^1 + \xi_i^2) \\ y_i - \left(w^T \phi_Q(x_i) - b\right) \leq \varepsilon + \xi_i^1, \qquad i = 1, \ldots, n \\ s.\text{à} \quad w^T \phi_Q(x_i) + b - y_i \quad \leq \varepsilon + \xi_i^2, \qquad i = 1, \ldots, n \\ \xi_i^1 \geq 0, \xi_i^2 \geq 0, (w, b) \in \mathbb{R}^d \times \mathbb{R} \end{cases}$$

$\{(x_i, y_i\}_{i=1}^{n}$ where $x_i \in \mathbb{R}^d, y_i \in \mathbb{R}$

- Linear regression $y = w^T x + b$ if $d = 2 \rightarrow x = (x^1, x^2)$ and $w = (w_1, w_2)$
- Nolinear Regression: $d = 2 \rightarrow x = (x^1, x^2)$
- $\phi_2(x) = \begin{pmatrix} 1, \\ x_1, x_2, \\ x_1^2, x_1 x_2, x_2^2 \end{pmatrix}$
- $y = w_1 x^1 + w_2 x^2 + w_3 x^1 x^2 + w_4 (x^1)^2 + w_5 (x^2)^2 + b$
- $\rightarrow y = w^T \ \phi_2(x) + b \rightarrow w = (w_1, w_2, w_3, w_4, w_5)$

ε-band SVRNR : Primal Algorithm

- Input: training set : $\{(x_i, y_i\}_{i=1}^{n}$ where $x_i \in \mathbb{R}^d, y_i \in \mathbb{R}$
- Choose an **appropriate map ϕ_Q**, parameter $\varepsilon > 0$ and penalty parameter $C > 0$
- Contruct and solve the optimzation problem $\mathrm{PLS} - \varepsilon\text{–band} - SVR\mathrm{NR}$ obtaining (w^*, b^*)

$$
\mathrm{PLS} - \varepsilon\text{–band} - SVR\mathrm{NR}: \begin{cases} \displaystyle min \quad \frac{1}{2} w^T w + C \sum_{i=1}^{n} (\xi_i^1 + \xi_i^2) & \\ \quad y_i - \left(w^T \phi_Q(x_i) - b\right) \leq \varepsilon + \xi_i^1, & i = 1, \dots, n \\ s.à \quad w^T \phi_Q(x_i) + b - y_i \quad \leq \varepsilon + \xi_i^2, & i = 1, \dots, n \\ \quad \xi_i^1 \geq 0, \xi_i^2 \geq 0, (w, b) \in \mathbb{R}^d \times \mathbb{R} & \end{cases}
$$

- Contruit the separating hyperplane $(w^*)^T \phi_Q(x) + b^*$ and the decision function is

$$
h(x) = (w^*)^T \phi_Q(x) + b^*
$$

ε-band SVRNR : Dual Form

$$\text{DLS} - \varepsilon\text{–band} - SVR\text{NR}\begin{cases} min \quad \dfrac{1}{2}\displaystyle\sum_{i,j=1}^{n}(\mu_i - \lambda_i)(\mu_j - \lambda_j)(\boldsymbol{\phi_Q}(\boldsymbol{x_i})^{\boldsymbol{T}}\boldsymbol{\phi_Q}(\boldsymbol{x_j})) - \varepsilon\sum_{i=1}^{n}(\mu_i + \lambda_i) + \sum_{i=1}^{n} y_i(\mu_i - \lambda_i) \\ s.\text{à} \qquad \displaystyle\sum_{i=1}^{n}(\mu_i - \lambda_i) = 0 \\ \qquad C \geq \lambda_i \geq 0, C \geq \mu_i \geq 0, \qquad i = 1, \ldots, n \end{cases}$$

ε-band SVRNR : Dual Algorithm

- Input: training set : $\{(x_i, y_i\}_{i=1}^{n}$ where $x_i \in \mathbb{R}^d, y_i \in \mathbb{R}$
- Choose an appropriate parameter $\varepsilon > 0$, map ϕ_Q and penalty parameter $C > 0$
- Contruct and solve the optimzation problem $\mathrm{DLS} - \varepsilon\mathrm{-band} - SVR$ obtaining $\lambda^* = (\lambda_1^*, \ldots, \lambda_n^*)$ and $\mu^* = (\mu_1^*, \ldots, \mu_n^*)$

$$\mathrm{DLS} - \varepsilon\mathrm{-band} - SVR\mathrm{NR} \begin{cases} min \quad \frac{1}{2} \sum_{i,j=1}^{n} (\mu_i - \lambda_i)(\mu_j - \lambda_j) \boldsymbol{(\phi_Q(x_j)^T \phi_Q(x_i))} - \varepsilon \sum_{i=1}^{n} (\mu_i + \lambda_i) + \sum_{i=1}^{n} y_i (\mu_i - \lambda_i) \\ s.\text{à} \qquad \sum_{i=1}^{n} (\mu_i - \lambda_i) = 0 \\ \qquad C \geq \lambda_i \geq 0, C \geq \mu_i \geq 0, \qquad i = 1, \ldots, n \end{cases}$$

if $\lambda^* \neq 0$ and $\mu^* \neq 0$, the solution to the problem, (w^*, b^*), can be obtained in the following way

- $w^* = \sum_{i=1}^{n} (\mu_i - \lambda_i) \phi_Q(x_i)$,
- for any component $\lambda_j^* \in]0, C[$ of λ^*, $b^* = y_j - (w^*)^T \phi_Q(x_j) + \varepsilon$
- Or for any component $\mu_j^* \in]0, C[$ of μ^*, $b^* = y_j - (w^*)^T \phi_Q(x_j) - \varepsilon$

$$h(x) = (w^*)^T \phi_Q(x) + b^*$$