CZ4041/SC4000: Machine Learning

Additional Notes: Kernel Regression

Li Boyang, Albert
School of Computer Science and Engineering,
NTU, Singapore

Acknowledgements: Slides are modified from the version prepared by Dr. Sinno Pan.

Regularized Linear Regression with Kernels

- Primal objective

$$\mathcal{T}(\boldsymbol{w}) = \frac{1}{2}\sum_{i=1}^{N}(\boldsymbol{w}\cdot\phi(\boldsymbol{x}_i) - y_i)^2 + \frac{\lambda}{2}\|\boldsymbol{w}\|_2^2$$

- It can be reformulated in terms of a dual form where kernel function arises naturally

$$\frac{\partial\mathcal{T}(\boldsymbol{w})}{\partial\boldsymbol{w}} = 0 \Rightarrow \sum_{i=1}^{N}(\boldsymbol{w}\cdot\phi(\boldsymbol{x}_i) - y_i)\phi(\boldsymbol{x}_i) + \lambda\boldsymbol{w} = 0$$

Closed Form Solution

- By using kernel trick $k(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$

$$f(\boldsymbol{x}) = \boldsymbol{w} \cdot \boldsymbol{\phi}(\boldsymbol{x})$$

$$\mathbf{k}_{\boldsymbol{i}}(\boldsymbol{x}) = k(\boldsymbol{x}_{\boldsymbol{i}}, \boldsymbol{x})$$

$$\boldsymbol{f}(\boldsymbol{x}) = \boxed{\mathbf{k}(\boldsymbol{x})}^{\boldsymbol{T}} (\boxed{\mathbf{K}} + \lambda \mathbf{I})^{-1} \boldsymbol{y}$$

$$\mathbf{K} = \begin{pmatrix} k(\boldsymbol{x}_1, \boldsymbol{x}_1) & \cdots & k(\boldsymbol{x}_1, \boldsymbol{x}_N) \\ \vdots & \ddots & \vdots \\ k(\boldsymbol{x}_N, \boldsymbol{x}_1) & \cdots & k(\boldsymbol{x}_N, \boldsymbol{x}_N) \end{pmatrix}$$

$$= \begin{pmatrix} \phi(\boldsymbol{x}_1) \cdot \phi(\boldsymbol{x}_1) & \cdots & \phi(\boldsymbol{x}_1) \cdot \phi(\boldsymbol{x}_N) \\ \vdots & \ddots & \vdots \\ \phi(\boldsymbol{x}_N) \cdot \phi(\boldsymbol{x}_1) & \cdots & \phi(\boldsymbol{x}_N) \cdot \phi(\boldsymbol{x}_N) \end{pmatrix}$$

Induction

$$\sum_{i=1}^{N}(\boldsymbol{w}\cdot\phi(\boldsymbol{x}_i)-y_i)\phi(\boldsymbol{x}_i)+\lambda\boldsymbol{w}=0$$

$$\Downarrow$$

$$\boldsymbol{w}=-\frac{1}{\lambda}\sum_{i=1}^{N}(\boldsymbol{w}\cdot\phi(\boldsymbol{x}_i)-y_i)\phi(\boldsymbol{x}_i)$$

$$=\sum_{i=1}^{N}a_i\phi(\boldsymbol{x}_i) \qquad a_i=-\frac{1}{\lambda}(\boldsymbol{w}\cdot\phi(\boldsymbol{x}_i)-y_i)$$

Induction (cont.)

Denote by $\mathbf{\Phi} = \left(\phi(\boldsymbol{x}_1), \phi(\boldsymbol{x}_2), \dots, \phi(\boldsymbol{x}_N)\right)^T$

and $\boldsymbol{a} = (a_1, \dots, a_N)^T$

$$\boldsymbol{w} = \sum_{i=1}^{N} a_i \phi(\boldsymbol{x}_i) \Rightarrow \boldsymbol{w} = \mathbf{\Phi}^T \boldsymbol{a}$$

$$a_i = -\frac{1}{\lambda}(\boldsymbol{w} \cdot \phi(\boldsymbol{x}_i) - y_i) \Rightarrow \boldsymbol{a} = \frac{1}{\lambda}(\boldsymbol{y} - \mathbf{\Phi}\boldsymbol{w})$$

Induction (cont.)

$$\boldsymbol{w} = \boldsymbol{\Phi}^T \boldsymbol{a} \qquad \boldsymbol{a} = \frac{1}{\lambda}(\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{w})$$

$$\boldsymbol{a} = \frac{1}{\lambda}(\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{\Phi}^T \boldsymbol{a}) \Rightarrow \lambda \mathbf{I}\boldsymbol{a} = \boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{\Phi}^T \boldsymbol{a}$$

$$\left(\lambda \mathbf{I} + \boldsymbol{\Phi}\boldsymbol{\Phi}^T\right)\boldsymbol{a} = \boldsymbol{y}$$

$$\boldsymbol{a} = \left(\lambda \mathbf{I} + \boldsymbol{\Phi}\boldsymbol{\Phi}^T\right)^{-1} \boldsymbol{y}$$

Induction (cont.)

Kernel matrix
(Gram matrix)

$$\mathbf{K} = \begin{pmatrix} k(\boldsymbol{x}_1, \boldsymbol{x}_1) & \cdots & k(\boldsymbol{x}_1, \boldsymbol{x}_N) \\ \vdots & \ddots & \vdots \\ k(\boldsymbol{x}_N, \boldsymbol{x}_1) & \cdots & k(\boldsymbol{x}_N, \boldsymbol{x}_N) \end{pmatrix}$$

$$= \begin{pmatrix} \phi(\boldsymbol{x}_1) \cdot \phi(\boldsymbol{x}_1) & \cdots & \phi(\boldsymbol{x}_1) \cdot \phi(\boldsymbol{x}_N) \\ \vdots & \ddots & \vdots \\ \phi(\boldsymbol{x}_N) \cdot \phi(\boldsymbol{x}_1) & \cdots & \phi(\boldsymbol{x}_N) \cdot \phi(\boldsymbol{x}_N) \end{pmatrix}$$

$$\boldsymbol{a} = \left(\lambda \mathbf{I} + \mathbf{\Phi}\mathbf{\Phi}^T\right)^{-1} \boldsymbol{y} \quad \Rightarrow \quad \boldsymbol{a} = (\lambda \mathbf{I} + \mathbf{K})^{-1} \boldsymbol{y}$$

$$\boldsymbol{w} = \mathbf{\Phi}^T \boldsymbol{a} \quad \Rightarrow \quad \boldsymbol{w} = \boxed{\mathbf{\Phi}^T} (\lambda \mathbf{I} + \mathbf{K})^{-1} \boldsymbol{y}$$

Induction (cont.)

$$\boldsymbol{w} = \boldsymbol{\Phi}^T(\lambda\mathbf{I} + \mathbf{K})^{-1}\boldsymbol{y} \qquad f(\boldsymbol{x}) = \boldsymbol{w} \cdot \phi(\boldsymbol{x}) = \phi(\boldsymbol{x})^T\boldsymbol{w}$$

$$f(\boldsymbol{x}) = \boxed{\phi(\boldsymbol{x})^T\boldsymbol{\Phi}^T}(\mathbf{K} + \lambda\mathbf{I})^{-1}\boldsymbol{y}$$

$$= \boxed{\boldsymbol{k}(\boldsymbol{x})}(\mathbf{K} + \lambda\mathbf{I})^{-1}\boldsymbol{y}$$

A vector of N dimensions, where $\boldsymbol{k}_i(\boldsymbol{x}) = k(x_i, \boldsymbol{x})$

Induction (cont.)

- Therefore

$$f(\boldsymbol{x}) = \boldsymbol{w} \cdot \boldsymbol{\phi}(\boldsymbol{x})$$

$$\mathbf{k}_{\boldsymbol{i}}(\boldsymbol{x}) = k(\boldsymbol{x}_{\boldsymbol{i}}, \boldsymbol{x})$$

$$\boldsymbol{f}(\boldsymbol{x}) = \boxed{\mathbf{k}(\boldsymbol{x})}^{\boldsymbol{T}}(\boxed{\mathbf{K}} + \lambda\mathbf{I})^{-1}\boldsymbol{y}$$

$$\mathbf{K} = \begin{pmatrix} k(\boldsymbol{x}_1, \boldsymbol{x}_1) & \cdots & k(\boldsymbol{x}_1, \boldsymbol{x}_N) \\ \vdots & \ddots & \vdots \\ k(\boldsymbol{x}_N, \boldsymbol{x}_1) & \cdots & k(\boldsymbol{x}_N, \boldsymbol{x}_N) \end{pmatrix}$$

$$= \begin{pmatrix} \phi(\boldsymbol{x}_1) \cdot \phi(\boldsymbol{x}_1) & \cdots & \phi(\boldsymbol{x}_1) \cdot \phi(\boldsymbol{x}_N) \\ \vdots & \ddots & \vdots \\ \phi(\boldsymbol{x}_N) \cdot \phi(\boldsymbol{x}_1) & \cdots & \phi(\boldsymbol{x}_N) \cdot \phi(\boldsymbol{x}_N) \end{pmatrix}$$

Thank you!