CS/ECE/ME 532
Matrix Methods in Machine Learning

Welcome!

Announcements

Mon	Tue	Wed	Thu	Fri
			04/25 **Act. 22**	04/26
04/29	04/30 **Act. 23 (due 12/05)** **Act. 24 (due 12/07)**	05/01 **Assignment 9***	05/02 **Act. 24 due** **Review**	05/03 **Assignment 10***
05/06	05/07 **Assessment 4** **@ 5:35pm**			

If you have a conflict, you must let me know TODAY!
(fill out “Request for taking Assessment 4 at a different time”)

* We will drop the *two* lowest Assignments (instead of only one)

Activity 22

Recall:

$\boldsymbol{w}$ depends on $\boldsymbol{x}_1, y_1, \boldsymbol{x}_2, \boldsymbol{y}_2 \ldots$

Kernel

$$\boxed{\widehat{y} = \phi(\boldsymbol{x})^T \boldsymbol{w} \rightarrow \widehat{y} = \sum_i \alpha_i K(\boldsymbol{x}, \boldsymbol{x}_i)}$$

High-dimensional feature vector

weighted sum of similarities between feature vector and each training point

$$\boldsymbol{w}^* = \arg\min_{\boldsymbol{w}} ||\boldsymbol{\Phi w} - \boldsymbol{y}|| + \lambda||\boldsymbol{w}||^2$$

$$\boldsymbol{w}^* = (\boldsymbol{\Phi}^T\boldsymbol{\Phi} + \lambda \boldsymbol{I})^{-1}\boldsymbol{\Phi}^T\boldsymbol{y}$$

Activity 21, problem 2

$$\boldsymbol{w}^* = \boldsymbol{\Phi}^T(\boldsymbol{\Phi}\boldsymbol{\Phi}^T + \lambda \boldsymbol{I})^{-1}\boldsymbol{y}$$

$$\widehat{y} = \phi(\boldsymbol{x})^T \boldsymbol{w}^* \rightarrow \widehat{y} = \sum_i \alpha_i K(\boldsymbol{x}, \boldsymbol{x}_i)$$

$$\boldsymbol{\alpha} = (\boldsymbol{\Phi}\boldsymbol{\Phi}^T + \lambda \boldsymbol{I})^{-1}\boldsymbol{y}$$

where $\boldsymbol{\Phi}\boldsymbol{\Phi}^T$ has i, j entry $K(\boldsymbol{x}_i, \boldsymbol{x}_j)$

No need to compute $\phi(\cdot)$ to compute $K(\cdot,\cdot)$ or $\widehat{y}$!

Some Insights:

If $\lambda \uparrow$, $\boldsymbol{w}$ / α $\downarrow$ and $\widehat{y}(\boldsymbol{x}) \downarrow$

$K(\boldsymbol{x}, \boldsymbol{x}_i)$		$\widehat{y}(\boldsymbol{x})$
Narrow	$\longrightarrow$	Rough
Broad	$\longrightarrow$	Smooth

Kernel SVMs

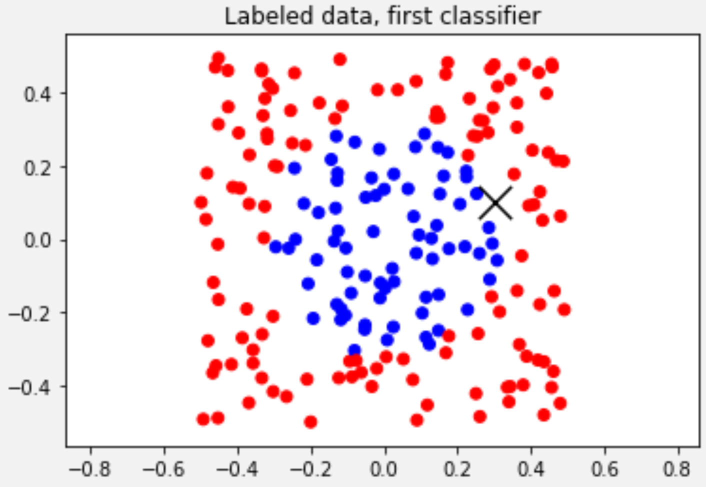
$$\widehat{y} = \text{sign}\left(\sum_i \alpha_i K(\boldsymbol{x}, \boldsymbol{x}_i)\right)$$

How do we find good α_i?

→ Iterative methods – subgradient descent

Support vectors have nonzero α_i

Kernel classification



$\boldsymbol{x} = \begin{bmatrix} 0.3 \\ 0.1 \end{bmatrix}$ How do we predict class of $\boldsymbol{x}$?

$$\widehat{y} = \text{sign}\left(\sum_i \alpha_i K(\boldsymbol{x}, \boldsymbol{x}_i)\right)$$

$$K(\boldsymbol{x}, \boldsymbol{x}_i) = \exp\left(-||\boldsymbol{x} - \boldsymbol{x}_i||^2\right)$$

$$\widehat{y} = \text{sign}\left(\sum_i \alpha_i \exp\left(-\left|\left|\begin{bmatrix} 0.3 \\ 0.1 \end{bmatrix} - \boldsymbol{x}_i\right|\right|^2\right)\right)$$