CS/ECE/ME 532
Matrix Methods in Machine Learning

Welcome!

The next three weeks…

- This week: finish Unit 5: hinge loss, SVM, stochastic GD
- Next week: start Unit 6: kernel methods
- The week after: finish Unit 6: neural networks + A4 review.
- Tuesday, May 7: Assessment 4

Due dates:

- Tomorrow: Assignment 8
- *Note*: two assignments are due in the last week of classes:
    - Wednesday, May 1: Assignment 9
    - Friday, May 3: Assignment 10

Activity 19

Loss functions

Classifying new data:

features *weights*

$\widehat{y} = \text{sign}(\boldsymbol{x}^T \boldsymbol{w})$

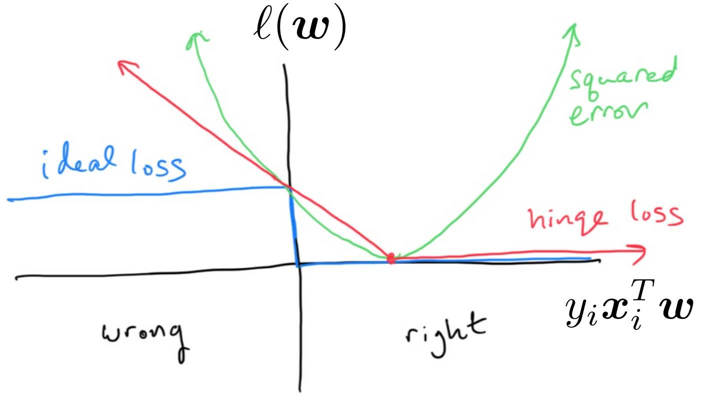
Training a classifier:

loss function

$\min_{\boldsymbol{w}} \ell(\boldsymbol{w}) + \lambda r(\boldsymbol{w})$

$\ell(\boldsymbol{w})$

- squared error $||\boldsymbol{X}\boldsymbol{w} - \boldsymbol{y}||_2^2$
- ideal (0-1) loss $\sum_i \frac{1}{2}|y_i - \text{sign}(\mathbf{x}_i^T \mathbf{w})|$
- $\boxed{\text{hinge loss } \sum_i (1 - y_i \boldsymbol{x}_i^T \boldsymbol{w})_+}$
- logistic loss $\log(1 + e^{-y_i \boldsymbol{x}_i^T \boldsymbol{w}})$

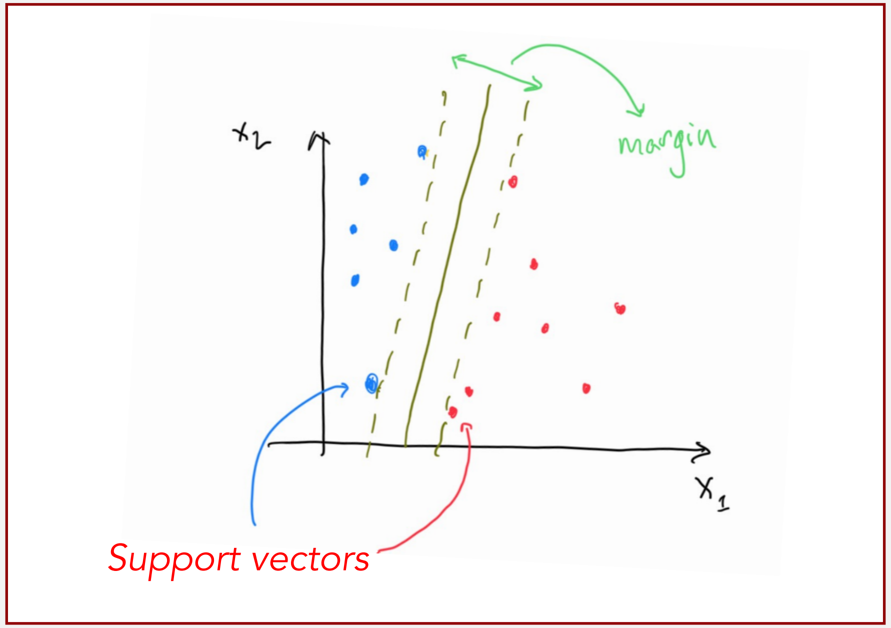


$y_i \boldsymbol{x}_i^T \boldsymbol{w}$

positive when correct

negative when wrong

Support Vector Machines



maximize margin
s.t. correct classification

$\longrightarrow$

minimize $||\tilde{\boldsymbol{w}}||^2$
s.t. $y_i \boldsymbol{x}_i^T \boldsymbol{w} \geq 1$ for $i = 1, \ldots$

For non-separable data: $\min_{\boldsymbol{w}} \sum_i (1 - y_i \boldsymbol{x}_i^T \boldsymbol{w})_+ + \lambda ||\tilde{\boldsymbol{w}}||_2^2$