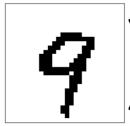
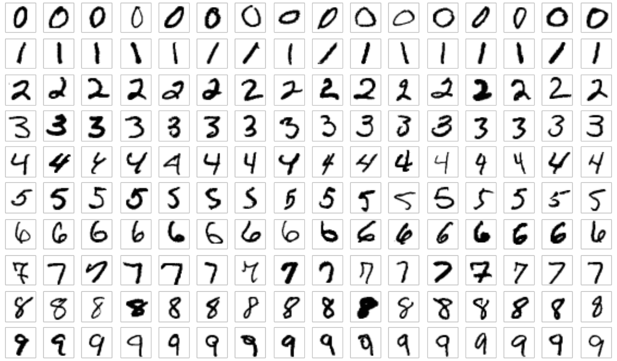
Activity 19: Hinge Loss and Support Vector Machines

Example: Handwritten digit recognition

$\boldsymbol{x} = \mathrm{vec}\Big($  $\Big)$ $\qquad \widehat{y} \approx \boldsymbol{x}^T \boldsymbol{w}$



Wikipedia MNIST dataset performance, April 2020

Type	Classifier	Error rate (%)
Linear classifier	Pairwise linear classifier	7.6[8]
Support-vector machine (SVM)	Virtual SVM, deg-9 poly, 2-pixel jittered	0.56[25]
Convolutional neural network	Committee of 20 CNNS with Squeeze-and-Excitation [29]	0.17[30]

Hinge Loss and Support Vector Machines

Classifying new data:

features

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

weights

Loss function

Training a classifier:

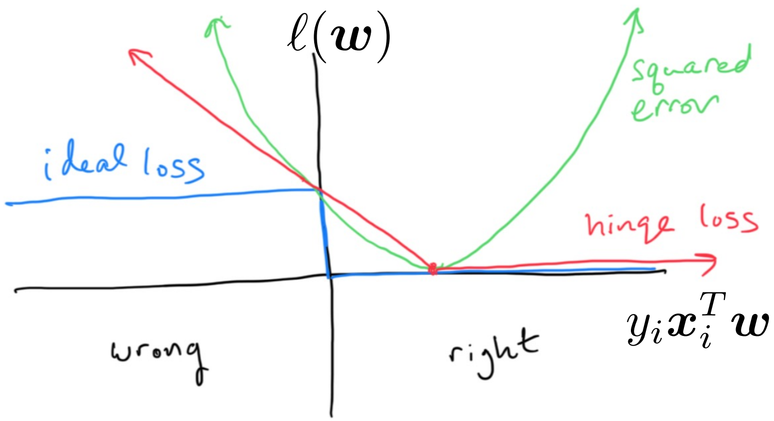
$$\min_{\boldsymbol{w}} \ell(\boldsymbol{w}) + \lambda r(\boldsymbol{w})$$

$\ell(\boldsymbol{w})$:

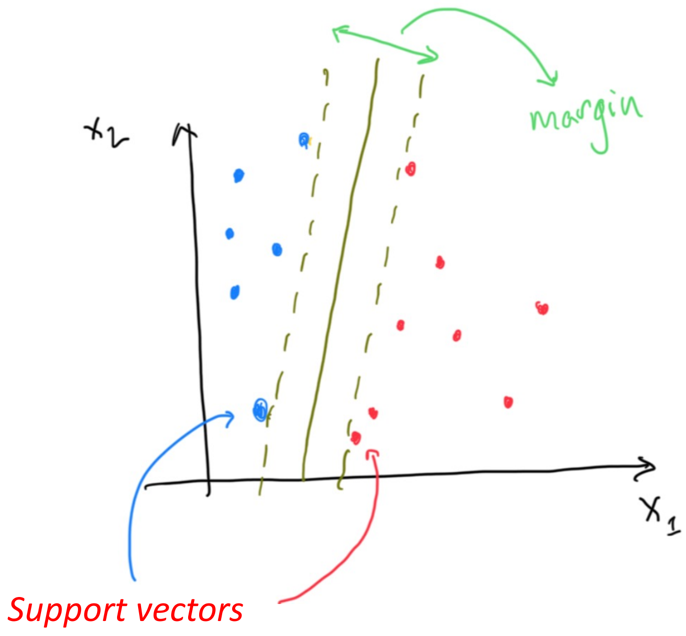
- squared error $||\boldsymbol{X}\boldsymbol{w} - \boldsymbol{y}||_2^2$
- ideal (0-1) loss $\sum_i \frac{1}{2}|y_i - \operatorname{sign}(\boldsymbol{x}_i^T \boldsymbol{w})|$
- hinge loss $\sum_i (1 - y_i \boldsymbol{x}_i^T \boldsymbol{w})_+$
- logistic loss $\log(1 + e^{-y_i \boldsymbol{x}_i^T \boldsymbol{w}})$

positive when correct classification,
negative when wrong

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



Support Vector Machines



maximize margin
s.t. correct classification

↓

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

↓

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