Machine Learning from Data

Lecture 8: Spring 2021

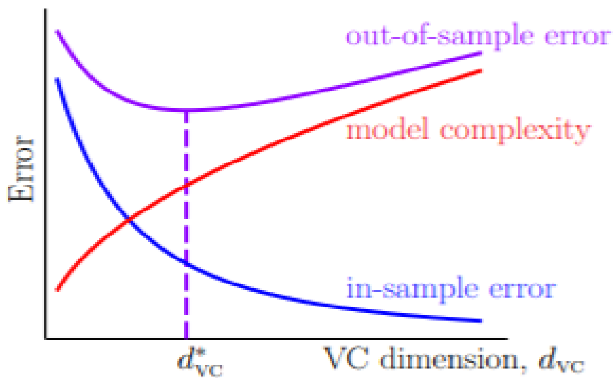
Today’s Lecture

- Linear Models
  - Classification and
  - Regression

VC Analysis

$$E_{\text{out}} \leq E_{\text{in}} + \Omega(d_{\text{VC}})$$

1. Did you fit your data well enough (E_{in})?
2. Are you confident your E_{in} will generalize to E_{out}



The VC Insuarance Co.

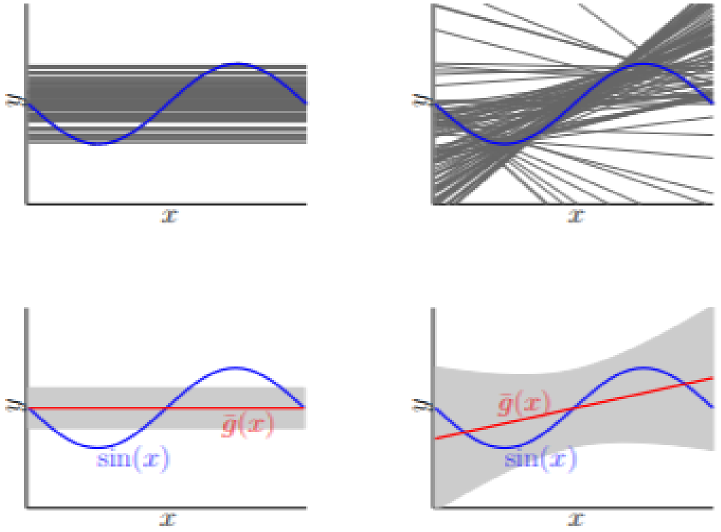
The VC warranty had conditions for becoming void:

You can't look at your data *before* choosing $\mathcal{H}$.
Data must be generated i.i.d from $P(\mathbf{x})$.
Data and test case from *same* $P(\mathbf{x})$ (same bin).

Bias-Variance Analysis

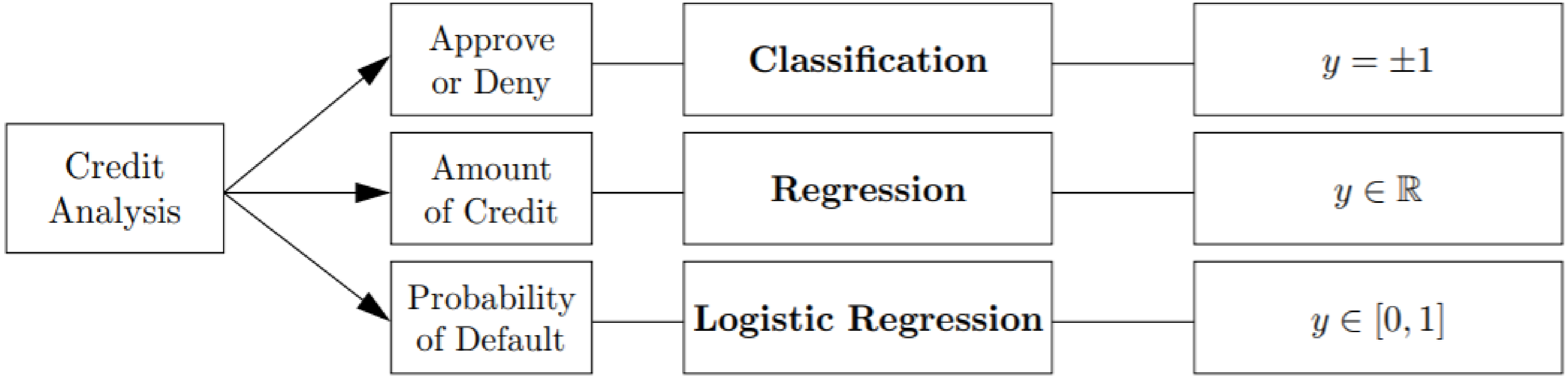
$$E_{\text{out}} = \textsf{bias} + \textsf{var}$$

1. How well can you fit your data (bias)?
2. How close to that best fit can you get (var)?



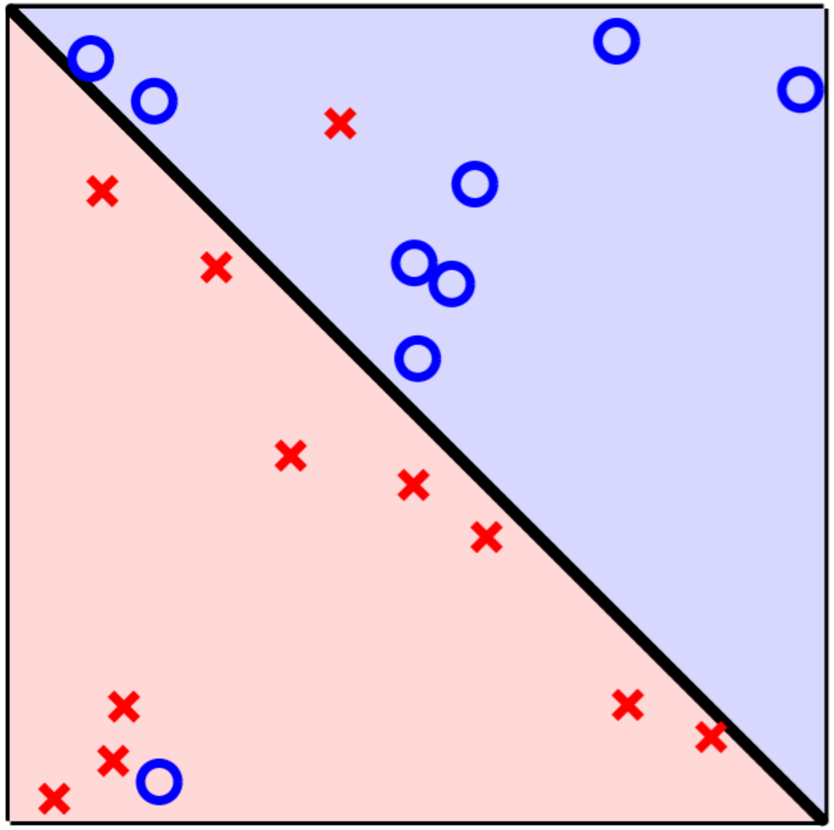
$\mathcal{H}_0$	$\mathcal{H}_1$
bias $= 0.50$;	bias $= 0.21$;
var $= 0.25$.	var $= 1.69$.
$E_{\text{out}} = 0.75$ ✓	$E_{\text{out}} = 1.90$

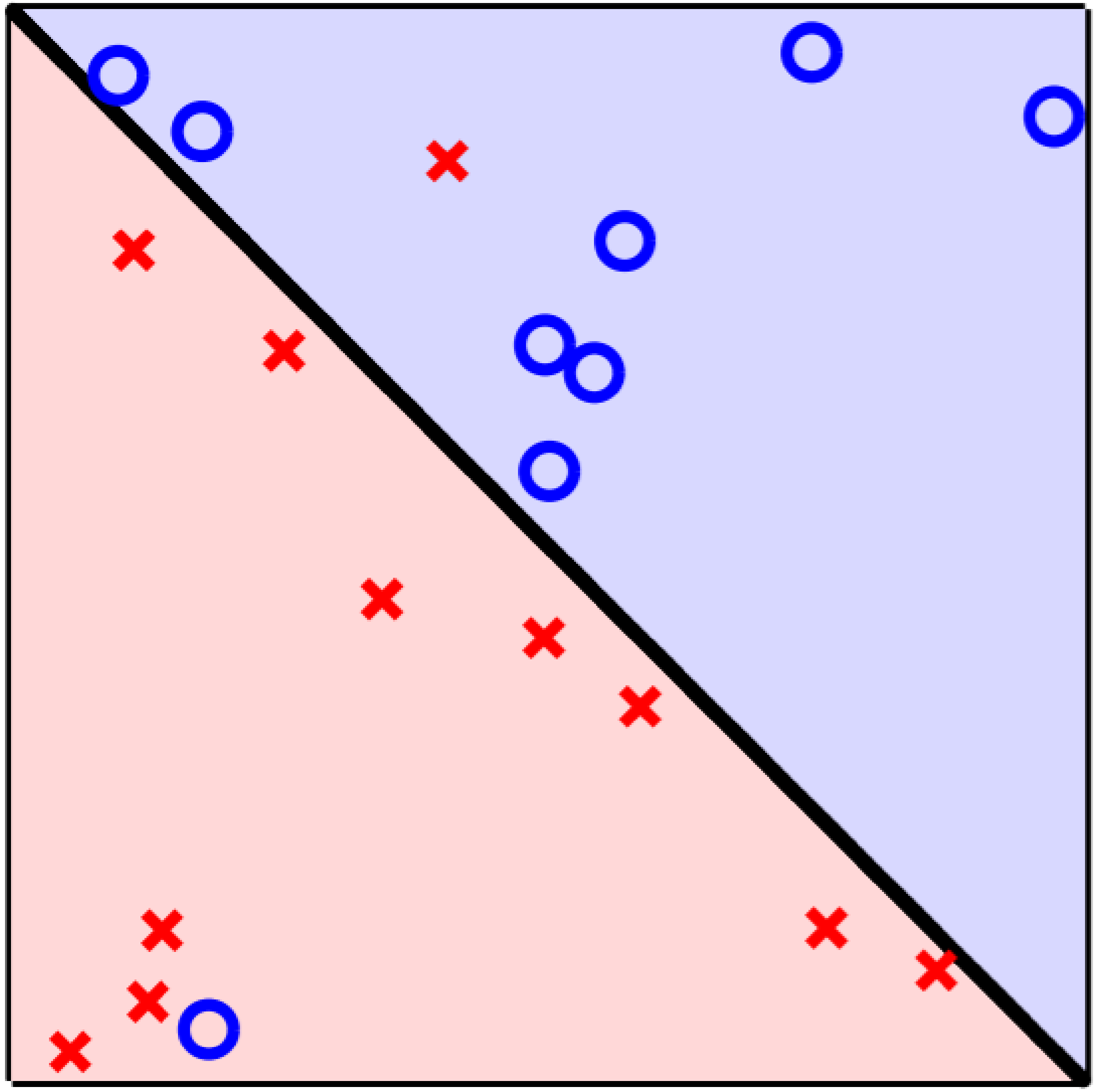
Linear Model



Classification

•





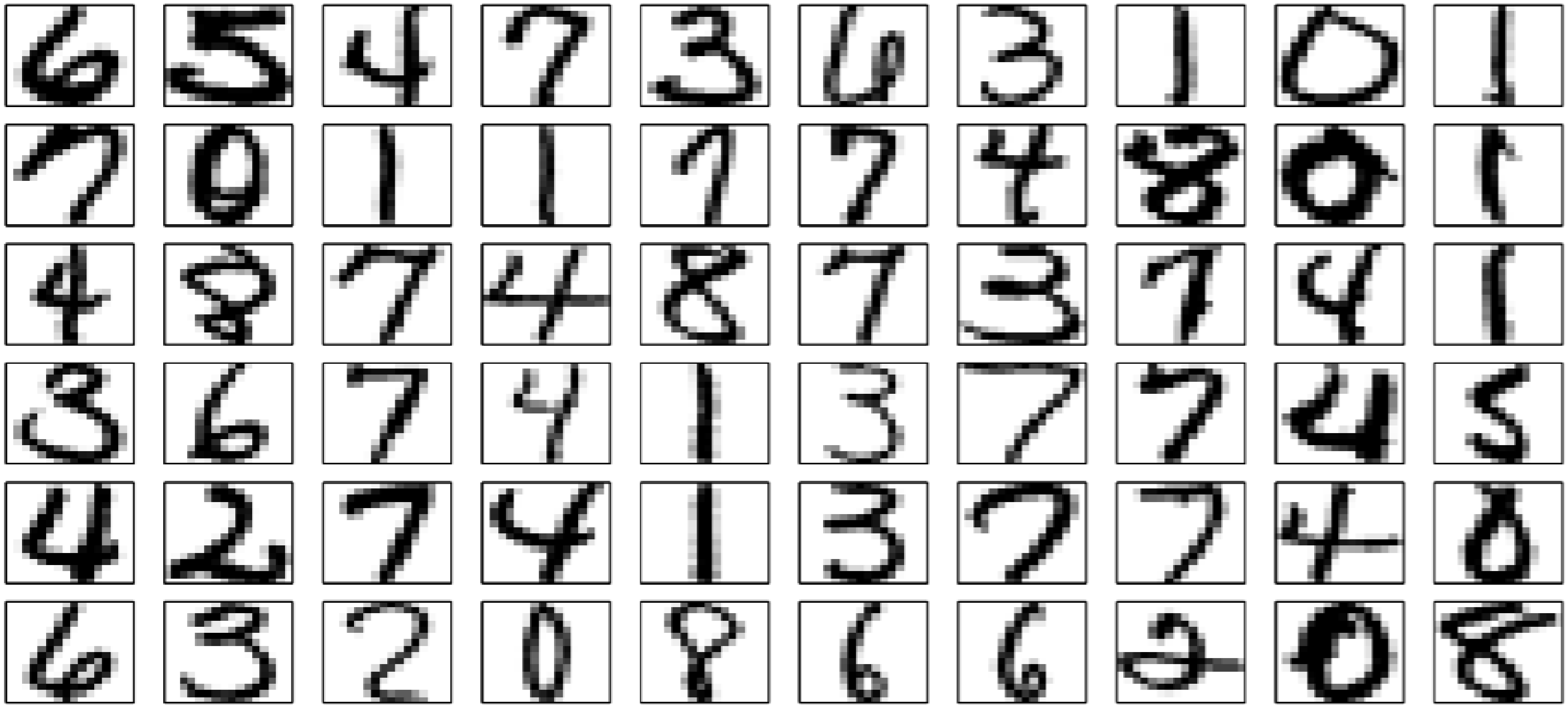
Minimizing E_{in} is a *hard* combinatorial problem.

The Pocket Algorithm

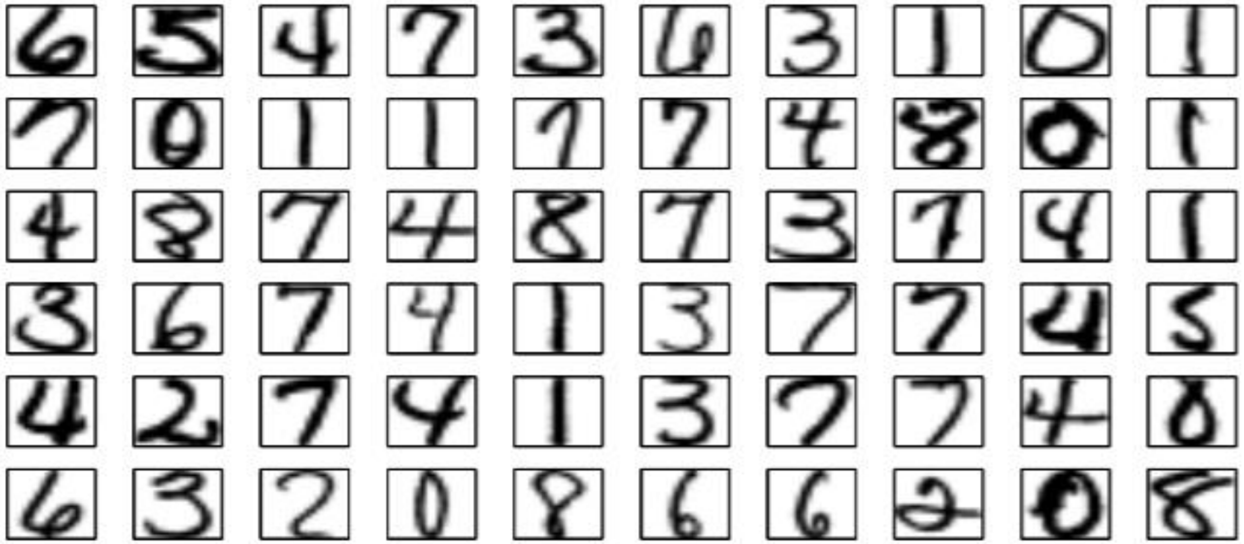
– Run PLA

– At each step keep the best E_{in} (and $\mathbf{w}$) so far.

(Its not rocket science, but it works.)



Each digit is a 16×16 image.



Each digit is a 16×16 image.

**[-1 -1 -1 -1 -1 -1 -1 -0.63 0.86 -0.17 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -0.99 0.3 1 0.31 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1
-1 -1 -0.41 1 0.99 -0.57 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -0.68 0.83 1 0.56 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -0.94 0.54
1 0.78 -0.72 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 0.1 1 0.92 -0.44 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -0.26 0.95 1 -0.16 -1 -1
-1 -0.99 -0.71 -0.83 -1 -1 -1 -1 -1 -0.8 0.91 1 0.3 -0.96 -1 -1 -0.55 0.49 1 0.88 0.09 -1 -1 -1 -1 0.28 1 0.88 -0.8 -1
-0.9 0.14 0.97 1 1 1 0.99 -0.74 -1 -1 -0.95 0.84 1 0.32 -1 -1 0.35 1 0.65 -0.10 -0.18 1 0.98 -0.72 -1 -1 -0.63 1 1
0.07 -0.92 0.11 0.96 0.30 -0.88 -1 -0.07 1 0.64 -0.99 -1 -1 -0.67 1 1 0.75 0.34 1 0.70 -0.94 -1 -1 0.54 1 0.02 -1 -1
-1 -0.90 0.79 1 1 1 1 0.53 0.18 0.81 0.83 0.97 0.86 -0.63 -1 -1 -1 -1 -0.45 0.82 1 1 1 1 1 1 1 0.13 -1 -1 -1 -1 -1
-1 -0.48 0.81 1 1 1 1 1 1 0.21 -0.94 -1 -1 -1 -1 -1 -1 -1 -0.97 -0.42 0.30 0.82 1 0.48 -0.47 -0.99 -1 -1 -1 -1]**

Feature Construction

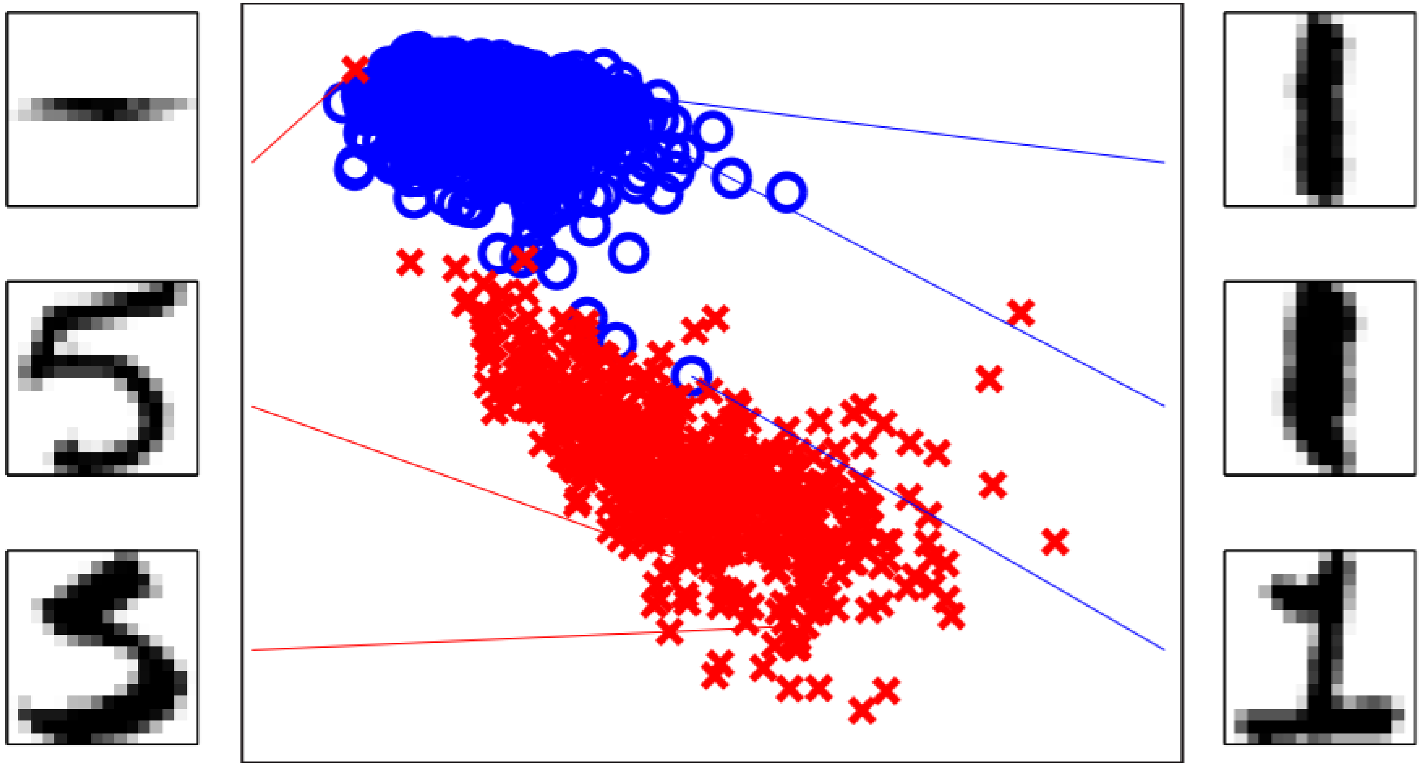
•

Linear Model

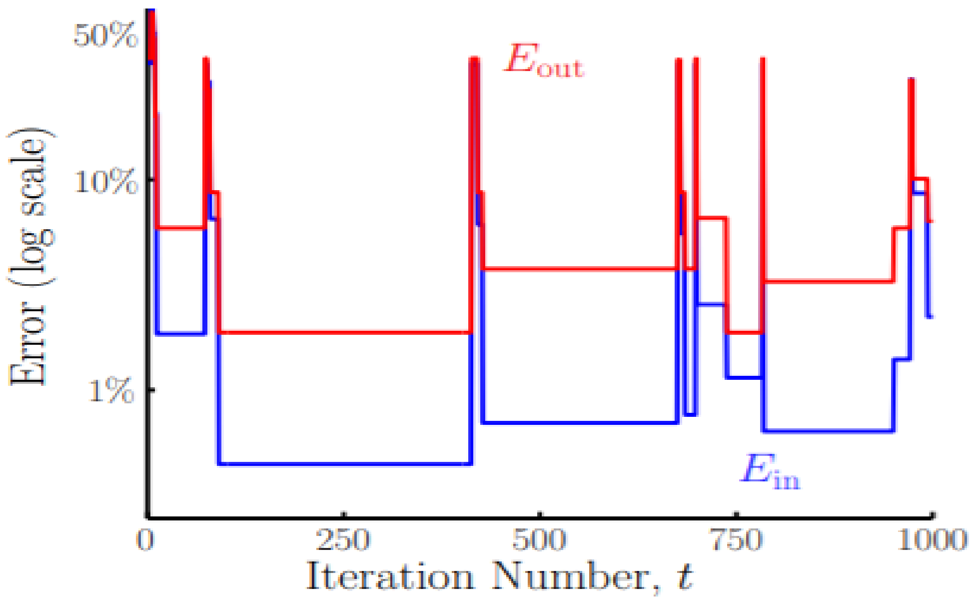
•

Intensity and Symmetry Features

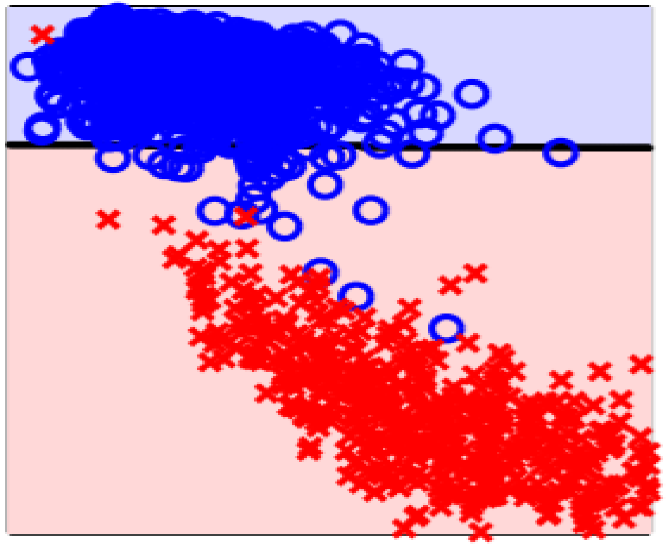
feature: an important property of the input that you think is useful for classification.
(dictionary.com: a prominent or conspicuous part or characteristic)

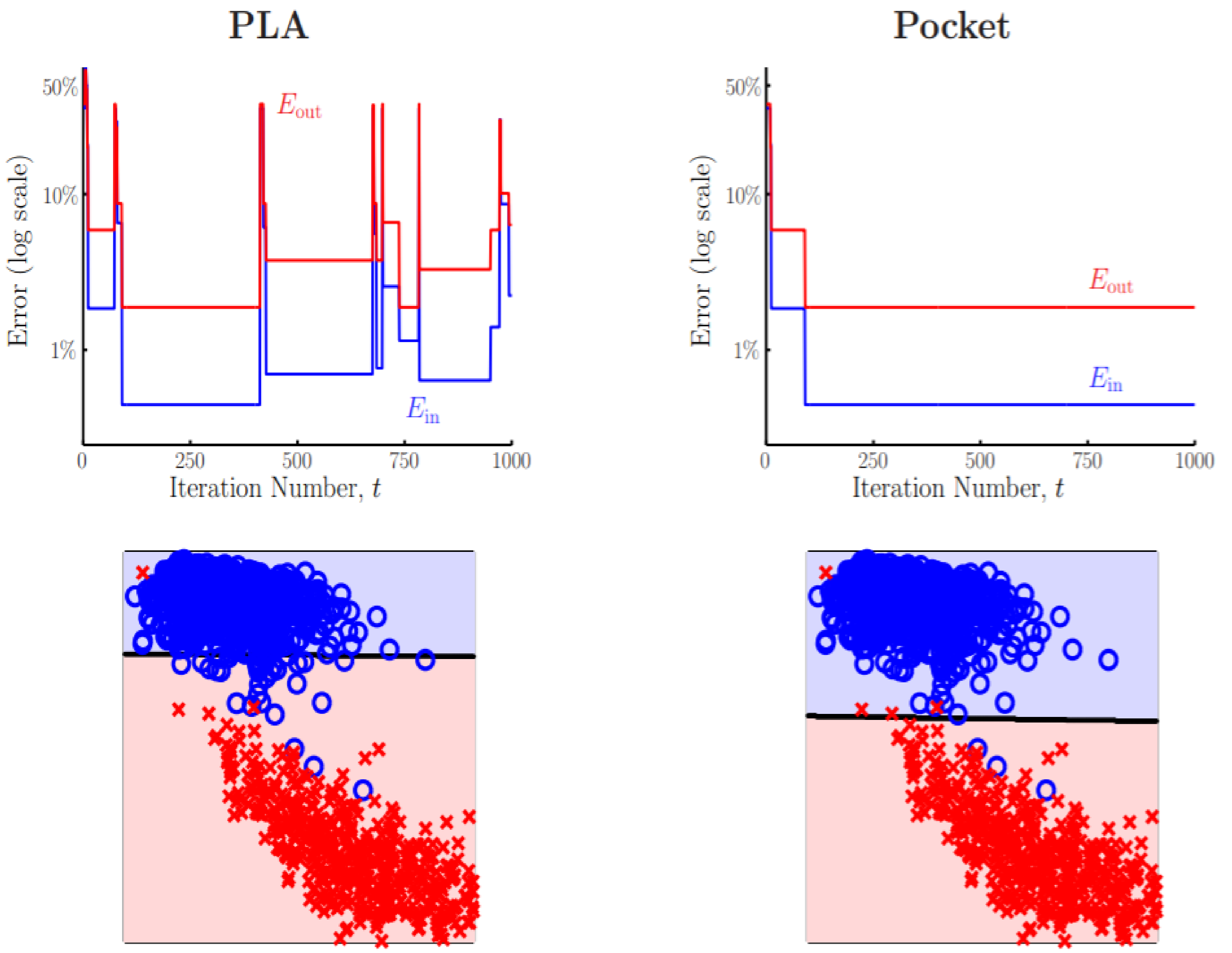


PLA



50%
10%
1%
Error (log scale)
E_{out}
E_{in}
0
250
500
750
1000
Iteration Number, t




PLA
Pocket
Error (log scale)
50%
10%
1%
E_{out}
E_{in}
0
250
500
750
1000
Iteration Number, t

Linear Regression

•

age	32 years
gender	male
salary	40,000
debt	26,000
years in job	1 year
years at home	3 years
...	...

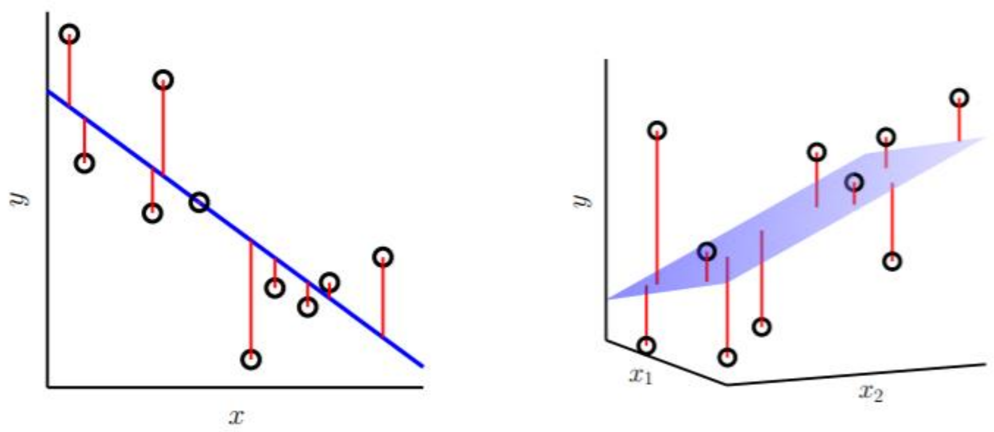
Classification: Approve/Deny

Regression: Credit Line (dollar amount)

regression $\equiv y \in \mathbb{R}$

$$h(\mathbf{x}) = \sum_{i=0}^{d} w_i x_i = \mathbf{w}^{\mathrm{T}}\mathbf{x}$$

Least Squares Linear Regression



$$y = f(\mathbf{x}) + \epsilon \qquad \longleftarrow \text{ noisy target } P(y|\mathbf{x})$$

in-sample error $\quad E_{\text{in}}(h) = \frac{1}{N}\sum_{n=1}^{N}(h(\mathbf{x}_n) - y_n)^2$

out-of-sample error $\quad E_{\text{out}}(h) = \mathbb{E}_{\mathbf{x}}[(h(\mathbf{x}) - y)^2]$

$$\left.\begin{array}{l} E_{\text{in}}(h) \\ E_{\text{out}}(h) \end{array}\right\} \quad h(\mathbf{x}) = \mathbf{w}^{\mathrm{T}}\mathbf{x}$$

Thanks!