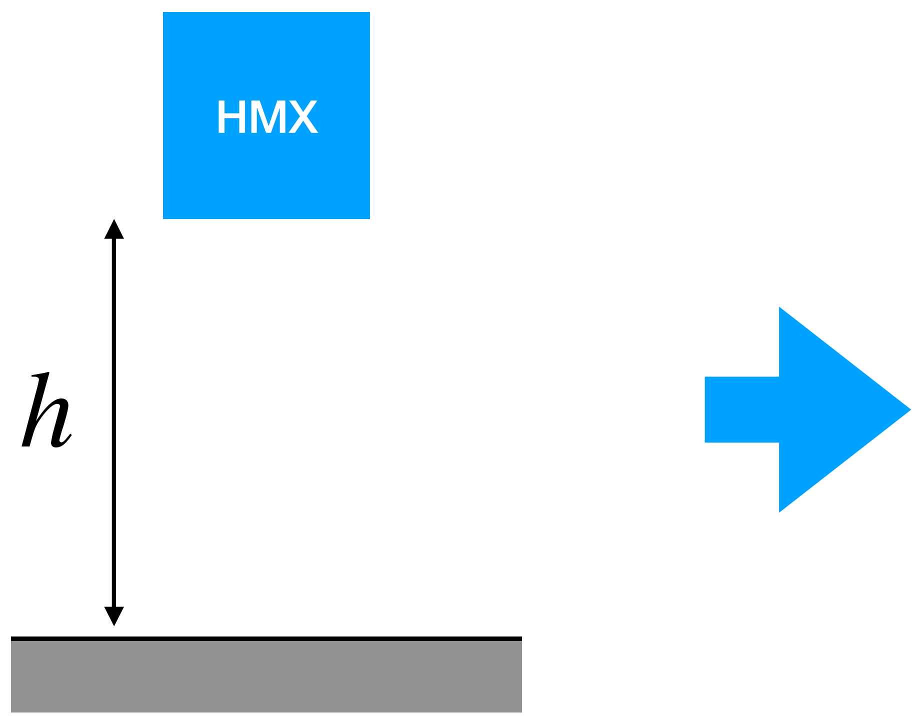
Lecture 16: Classification

Professor Ilias Bilionis

Logistic regression with one variable

PREDICTIVE
SCIENCE LABORATORY

Example: Sensitivity of energetic materials



Does it explode or not when you drop it from height h?

Experimental data

Height (cm)	**Results**
40.5	E E E E E E E E E E
36.0	E N E E E E N E E E
32.0	E E N E E E N E N E
28.5	N E N N E N N N E N
25.5	N N N N N N E N N N
22.5	N N N N N N N N N N

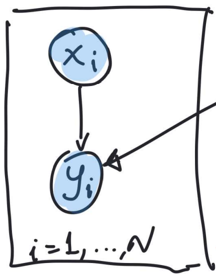
Data from L. Smith, "Los Alamos National Laboratory explosives orientation course: Sensitivity and sensitivity tests to impact, friction, spark and shock," Los Alamos National Lab, NM (USA), Tech. Rep., 1987

PREDICTIVE
SCIENCE LABORATORY

The logistic regression model

Given: $x_{1:N} = (x_1, \dots, x_N)$, $y_{1:N} = (y_1, \dots, y_N)$; $y_i \in \{0, 1\}$

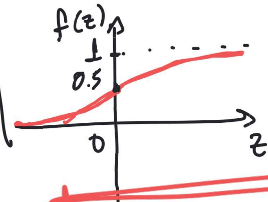
Find: $p(y \mid x, x_{1:N}, y_{1:N}) = ?$



Likelihood:

$$p(y_i = 1 \mid x_i, \underline{w}) = f(w_0 + w_1 x_i)$$

?



$$f(z) = \mathrm{sigm}(z) = \frac{\exp\{z\}}{1 + \exp\{z\}}$$

Important

$$\boxed{p(y_i = 1 \mid x_i, \underline{w}) = \mathrm{sigm}(w_0 + w_1 x_i)}$$

$$p(y_i = 0 \mid x_i, \underline{w}) = 1 - p(y_i = 1 \mid x_i, \underline{w})$$
$$= 1 - \mathrm{sigm}(w_0 + w_1 x_i)$$

$$p(y_i \mid x_i, \underline{w}) = \underbrace{\left[\mathrm{sigm}(w_0 + w_1 x_i)\right]^{y_i}}_{\text{activated when } y_i = 1} \cdot \underbrace{\left[1 - \mathrm{sigm}(w_0 + w_1 x_i)\right]^{1-y_i}}_{\text{activated when } y_i = 0}$$

$$p(y_{1:N} \mid x_{1:N}, \underline{w}) = \prod_{i=1}^{N} p(y_i \mid x_i, \underline{w})$$
$$= \prod_{i=1}^{N} \left[\mathrm{sigm}(w_0 + w_1 x_i)\right]^{y_i} \cdot \left[1 - \mathrm{sigm}(w_0 + w_1 x_i)\right]^{1-y_i}$$

PREDICTIVE SCIENCE LABORATORY

Training the model

$p(y_{1:N} | x_{1:N}, \underline{w})$; $\underline{w} \sim p(\underline{w})$

$$p(\underline{w} | x_{1:N}, y_{1:N}) \propto p(y_{1:N} | x_{1:N}, \underline{w})\, p(\underline{w})$$

NOT ANALYTICALLY AV.

Max. A Posteriori Estimate of $\underline{w}$:

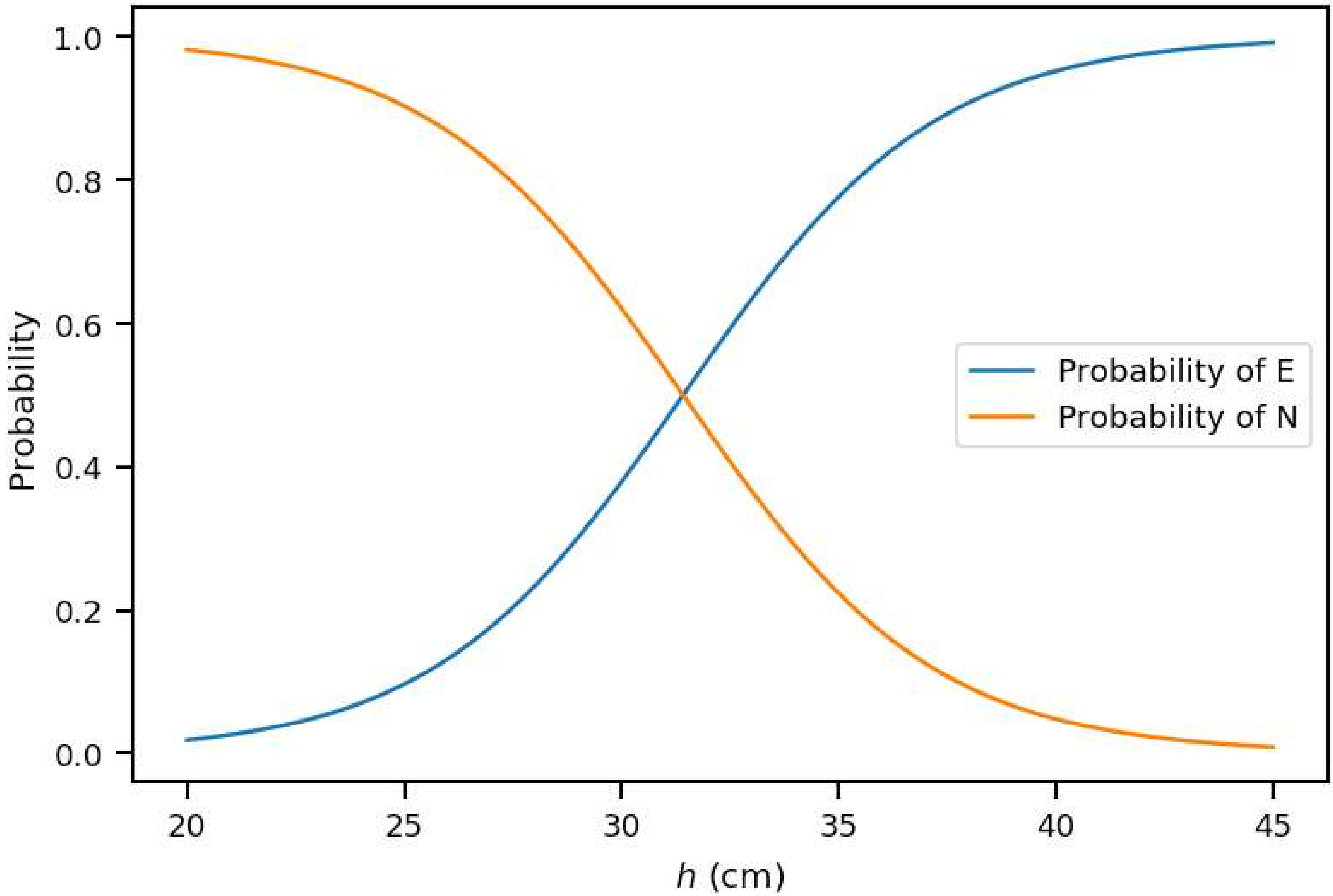
$$\max_{\underline{w}} \log p(y_{1:N} | x_{1:N}, \underline{w}) + \log p(\underline{w})$$

$$= \sum_{i=1}^{N} \log p(y_i | x_i, \underline{w}) + \cancel{\log p(\underline{w})} \to 0$$

$$= \sum_{i=1}^{N} \left\{ y_i \log \operatorname{sigm}(w_0 + w_1 x_i) + (1-y_i)\left[1 - \operatorname{sigm}(w_0 + w_1 x_i)\right] \right\}$$

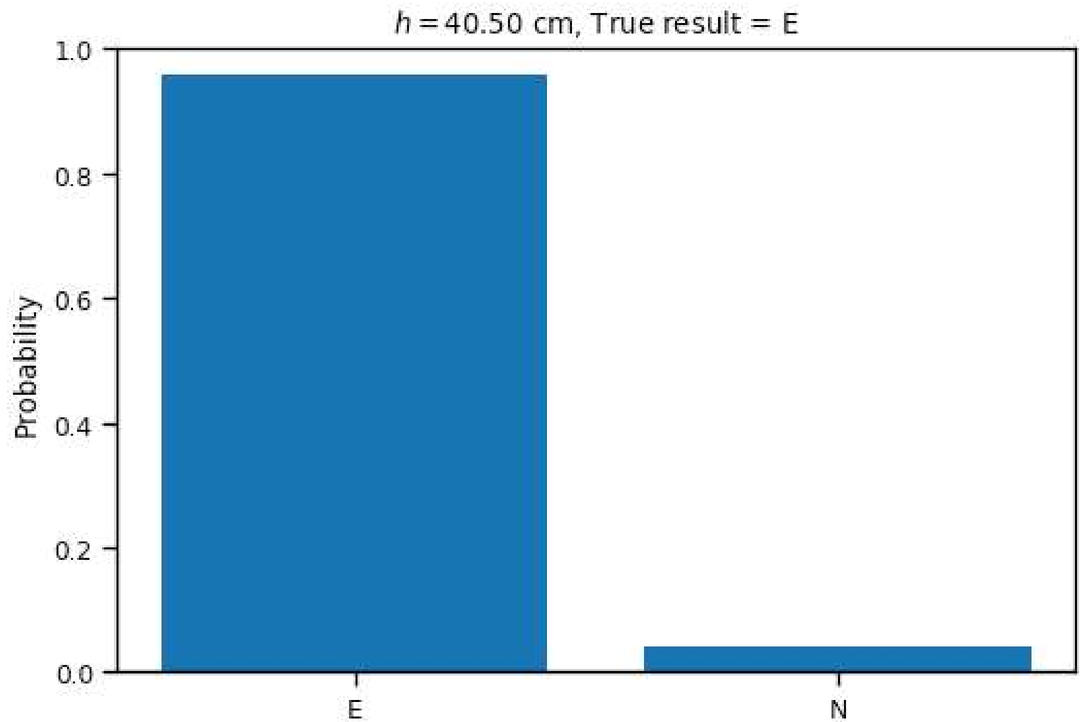
PREDICTIVE SCIENCE LABORATORY

How does the trained model look like?



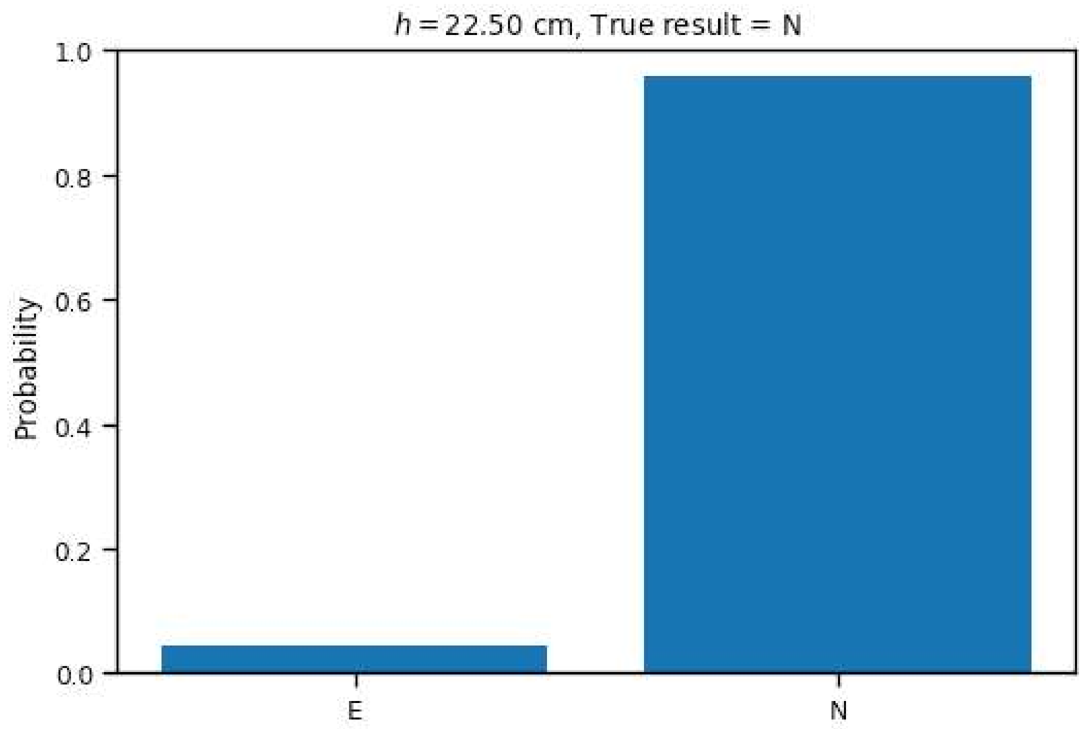
PREDICTIVE
SCIENCE LABORATORY

Making point-wise predictions



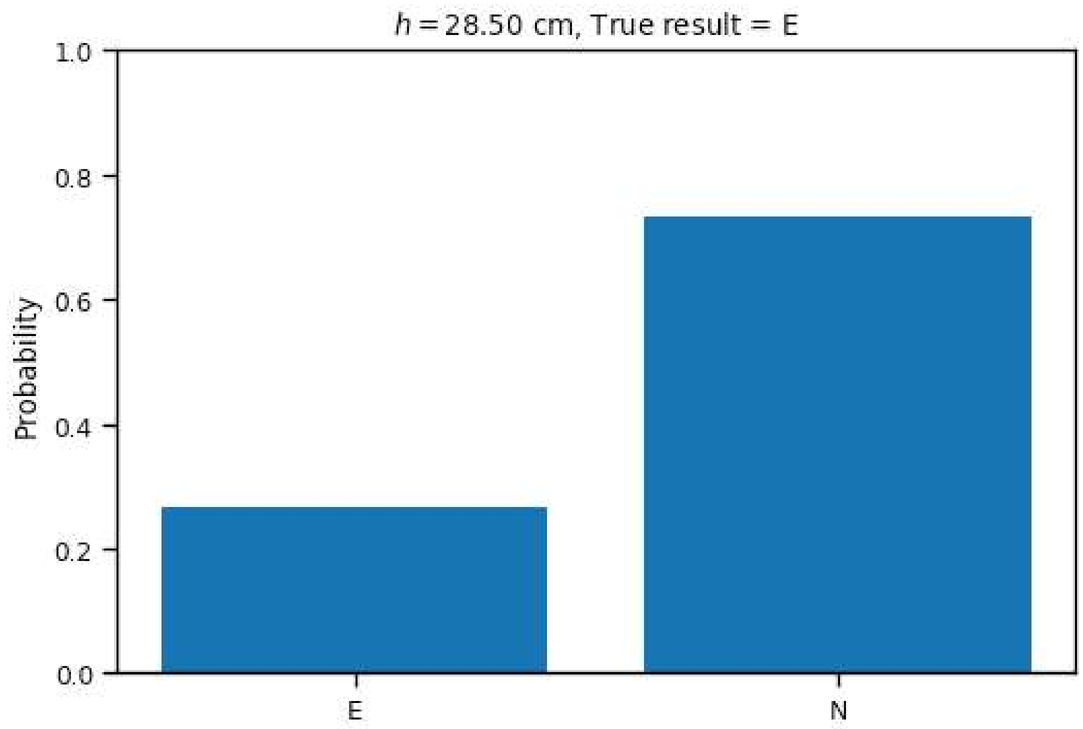
PREDICTIVE
SCIENCE LABORATORY

Making point-wise predictions



PREDICTIVE SCIENCE LABORATORY

Making point-wise predictions



PREDICTIVE
SCIENCE LABORATORY

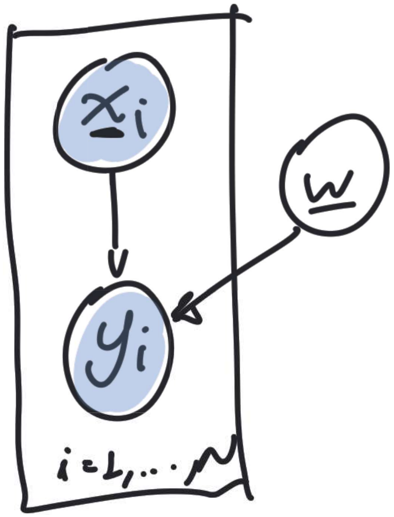
Lecture 16: Classification

Professor Ilias Bilionis

Logistic regression with many features

PREDICTIVE
SCIENCE LABORATORY

Combining logistic regression with generalized linear models



$$\underline{x}_i \in \mathbb{R}^D \;;\; \underline{x}_i = (x_{i1}, x_{i2}, \dots, x_{iD})$$

$$p(y_i = 1 \mid \underline{x}_i, \underline{w}) = \mathrm{sigm}(w_0 + w_1 x_{i1} + w_2 x_{i2} + \dots + w_D x_{iD})$$

$$\underline{w} = (w_0, w_1, \dots, w_D) \in \mathbb{R}^D .$$

$$\phi_1(\underline{x}), \dots, \phi_M(\underline{x})$$

Generalized Linear Model.

$$p(y_i = 1 \mid \underline{x}_i, \underline{w}) = \mathrm{sigm}\left(\sum_{j=1}^{M} w_j \phi_j(\underline{x}_i)\right)$$

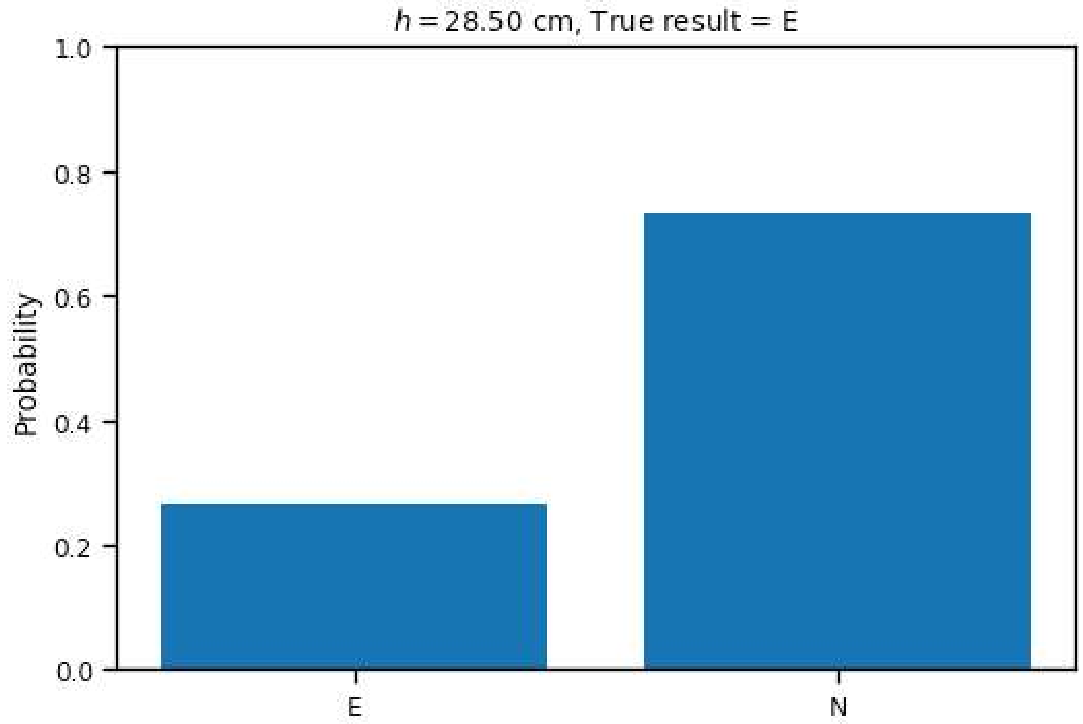
$$\underline{w} = (w_1, \dots, w_M) \in \mathbb{R}^M$$

Lecture 16: Classification

Professor Ilias Bilionis

Making decisions

HMX Example



How do you pick a single label?

PREDICTIVE
SCIENCE LABORATORY

Picking labels by minimizing the expected cost

$p(y \mid x, \underline{w})$

Pick label $\hat{y}$. Cost of choice $c(\hat{y}, y)$.

	True E	True N
Predict E	0	1
Predict N	100	0

$$\min_{\hat{y}} \mathbb{E}[c(\hat{y}, y) \mid x_{1:N}, y_{1:N}]$$

$$= \min_{\hat{y}} \int c(\hat{y}, y)\, \underline{p(y \mid x, x_{1:N}, y_{1:N})}\, dy$$

$$\simeq \min_{\hat{y}} \int c(\hat{y}, y)\, p(y \mid x, \underline{w})\, dy$$

Best decision when risk-neutral.

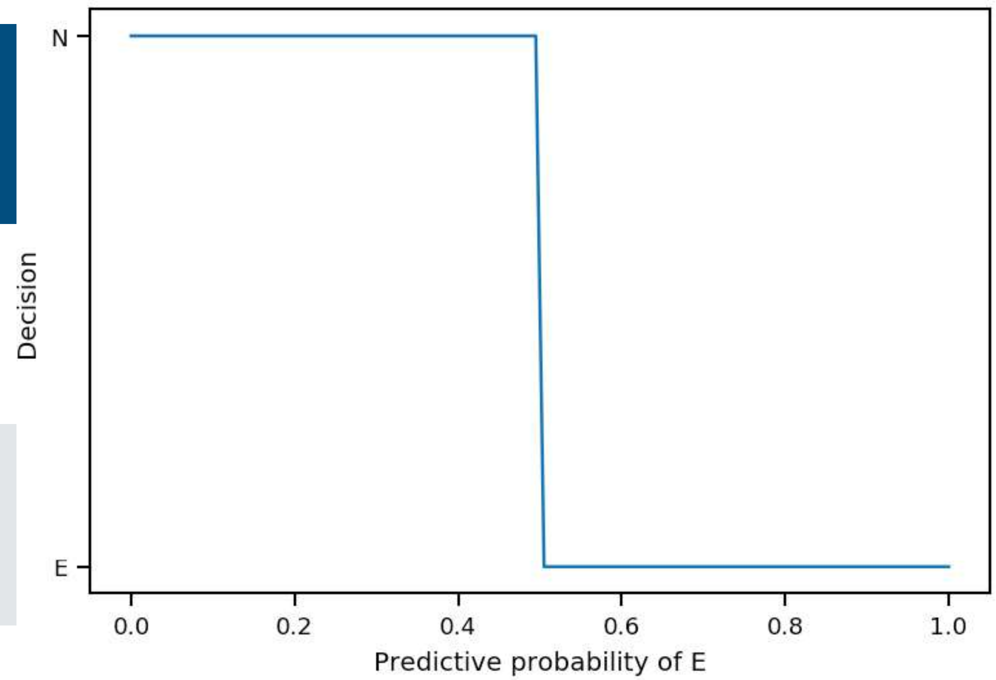
When risk-averse you need <u>utility theory</u>.

Jitesh Panchal Decision Making.

PREDICTIVE SCIENCE LABORATORY

The cost of making wrong predictions

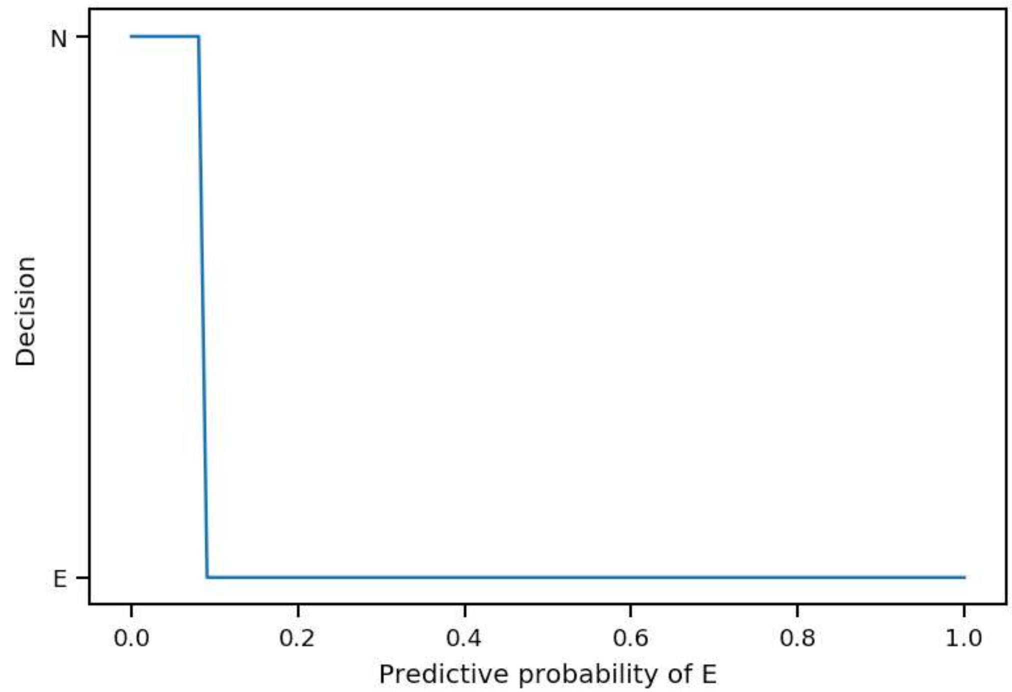
	True result E	True result N
Predict E	0	1
Predict N	1	0



PREDICTIVE
SCIENCE LABORATORY

The cost of making wrong predictions

	True result E	True result N
Predict E	0	1
Predict N	10	0



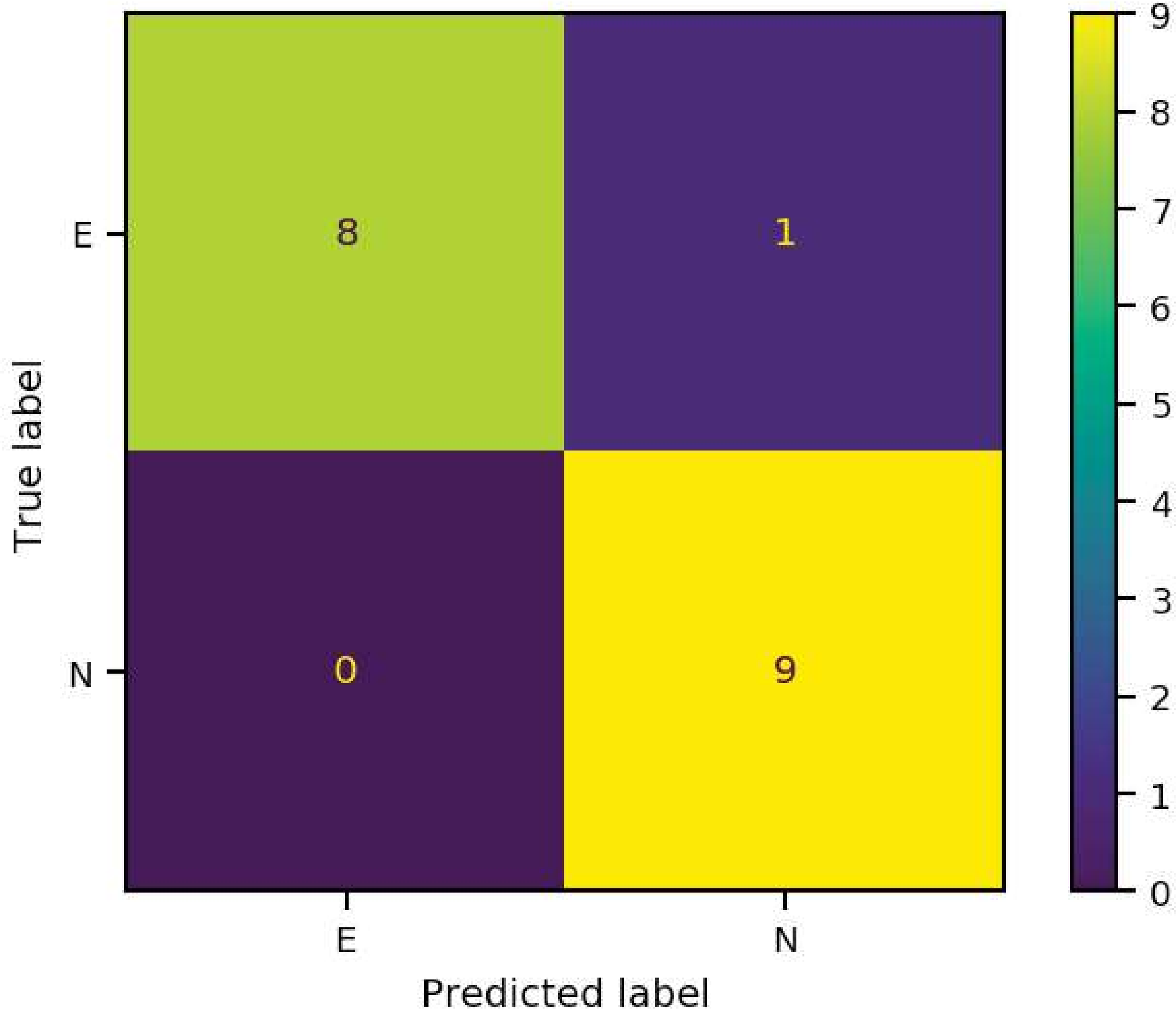
PREDICTIVE
SCIENCE LABORATORY

Lecture 16: Classification

Professor Ilias Bilionis

Diagnostics for classification

Confusion matrix



PREDICTIVE
SCIENCE LABORATORY

Accuracy score

obs $\begin{cases} x_1, x_2, \dots, x_{N^v} \\ y_1, y_2, \dots, y_{N^v} \end{cases}$

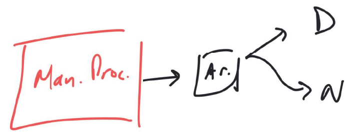
pred $\hat{y}_1, \hat{y}_2, \dots, \hat{y}_{N^v}$

$$\mathrm{acc}(\hat{y}_{1:N^v}, y_{1:N^v}) = \text{\% of observations correctly predicted}$$

$$= \frac{1}{N^v} \sum_{i=1}^{N^v} 1_{\{y_i\}}(\hat{y}_i)$$

PREDICTIVE
SCIENCE LABORATORY

Imbalanced data



N N N N N N D N N N . . .

$\text{Stupid Model}(x) = N$ with 100% prob.

↳ 99% accuracy because D happens only 1% of the time.

Because of imbalance between N and D.

True positives = TP = # of correctly predicted D.

True Negatives = TN = # of correctly predicted N.

False positives = FP = # of predicted D that were wrong.

False negatives = FN = # of predicted N that were wrong

$$\text{Sensitivity} = \frac{TP}{TP+FN}$$ = % of D that were predicted correctly.

$$\text{Specificity} = \frac{TN}{TN+FP}$$ = % of N that were predicted correctly

$$\text{balanced accuracy} = \frac{1}{2}(\text{Sensitivity} + \text{Specificity})$$

$$= \frac{1}{2}(\text{\% of corr. pred. D's} + \text{\% of corr. pred. N})$$

$$\text{Stupid Model's balan. acc.} = \frac{1}{2}(0+1) = 0.5$$

More Metrics

- Cross entropy loss
- Receiver operating characteristics curve
- f1-score
- Brier score
- …

PREDICTIVE
SCIENCE LABORATORY

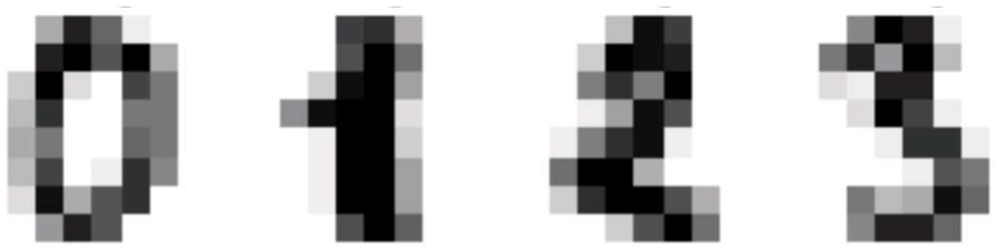
Lecture 16: Classification

Professor Ilias Bilionis

Multi-class logistic regression

PREDICTIVE
SCIENCE LABORATORY

Recognizing hand-written digits

inputs x =  **...**

labels y = **0** **1** **2** **3** **...**

Multi-class logistic regression model

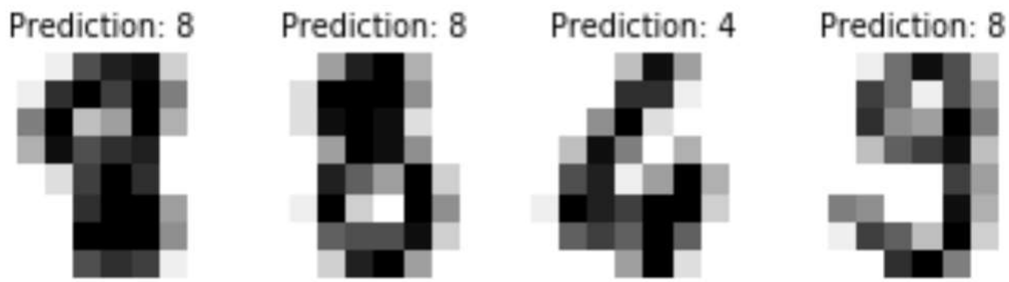
K different labels, $1, 2, \ldots, K$

$$p(y=k \mid \underline{x}, \underline{w}) = \frac{\exp\left\{\sum_{j=1}^{M} w_{jk}\,\varphi_j(\underline{x})\right\}}{\sum_{k'=1}^{K} \exp\left\{\sum_{j=1}^{M} w_{jk'}\,\varphi_j(\underline{x})\right\}}$$

(softmax transformation)

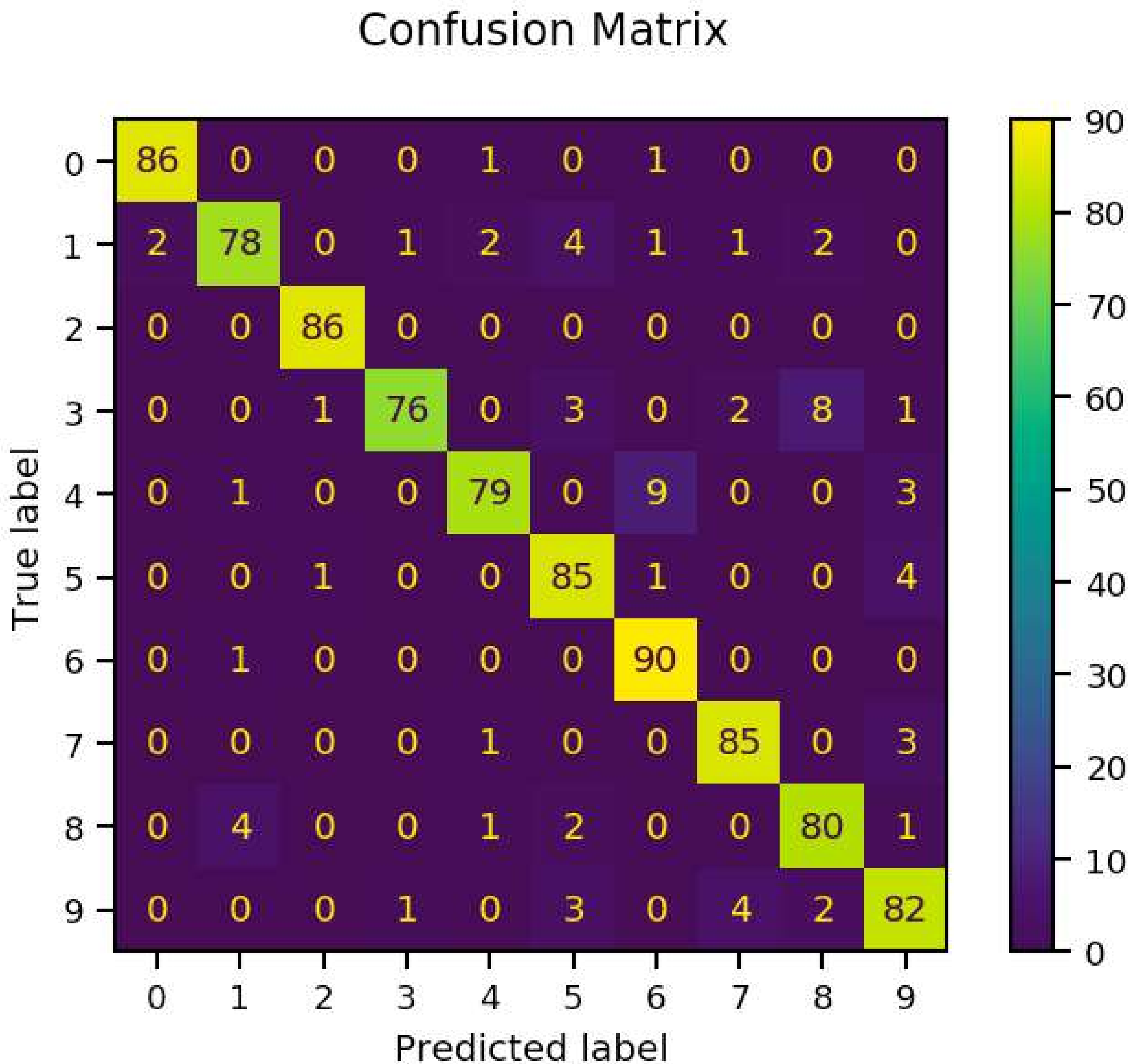
$$\underline{w} = (\underline{w}_1, \ldots, \underline{w}_K)$$

Results



PREDICTIVE
SCIENCE LABORATORY

Results



PREDICTIVE
SCIENCE LABORATORY