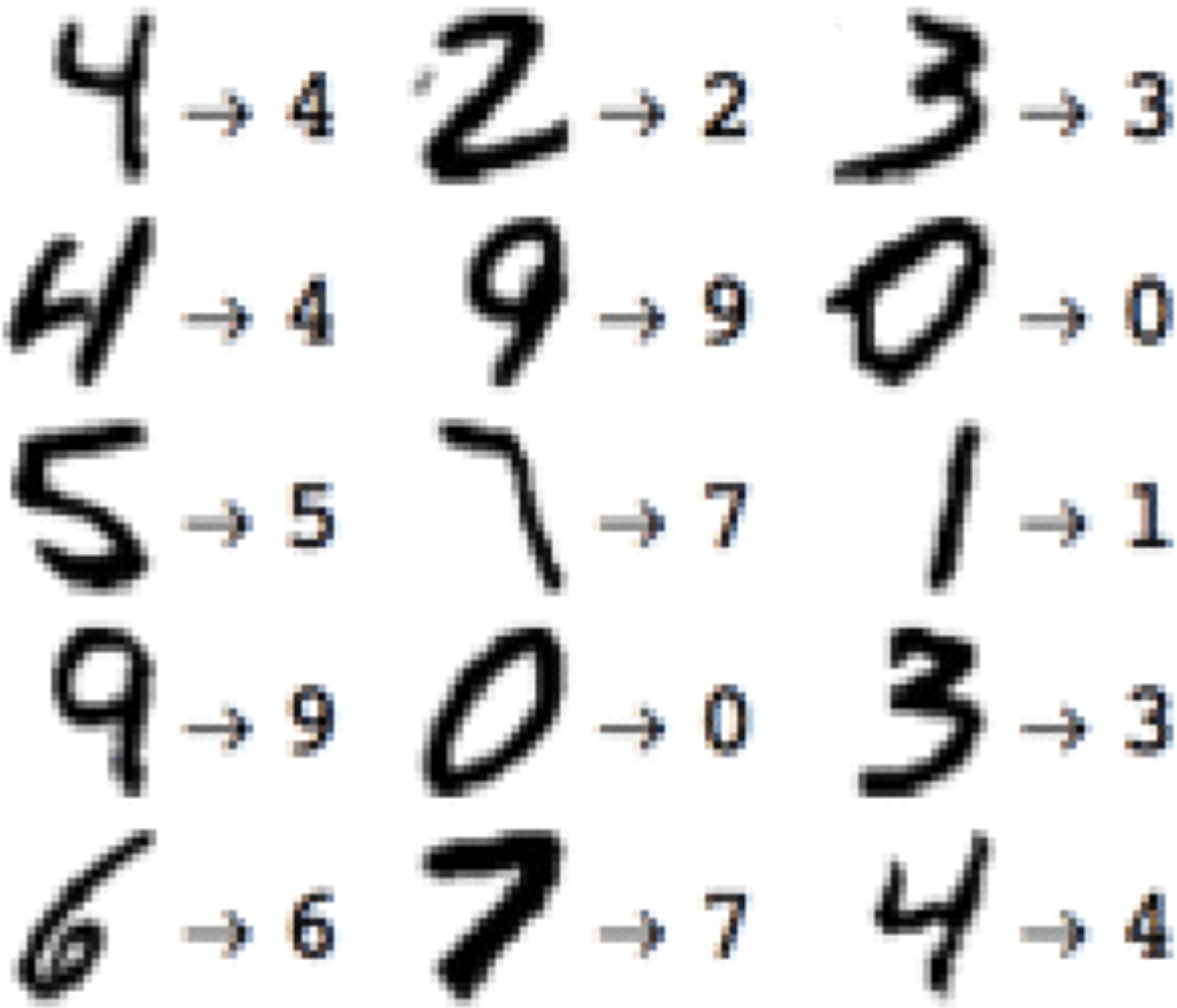
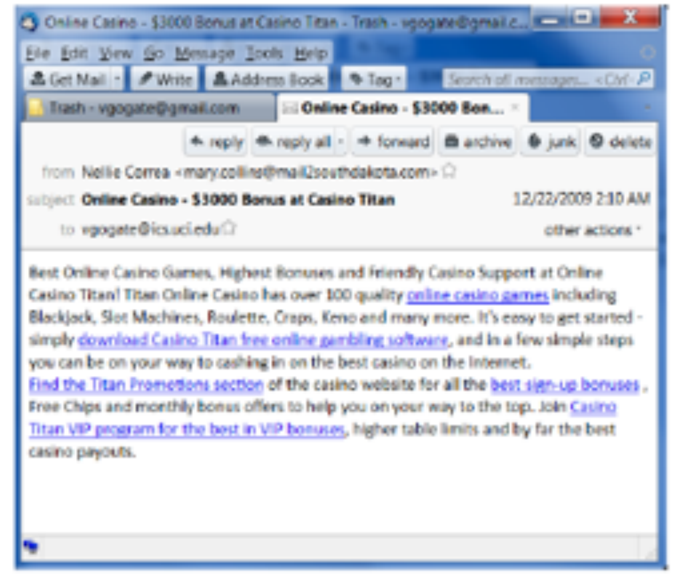
# Applied Data Science
# fall 2017
# Session 7: Classification. Logistic regression

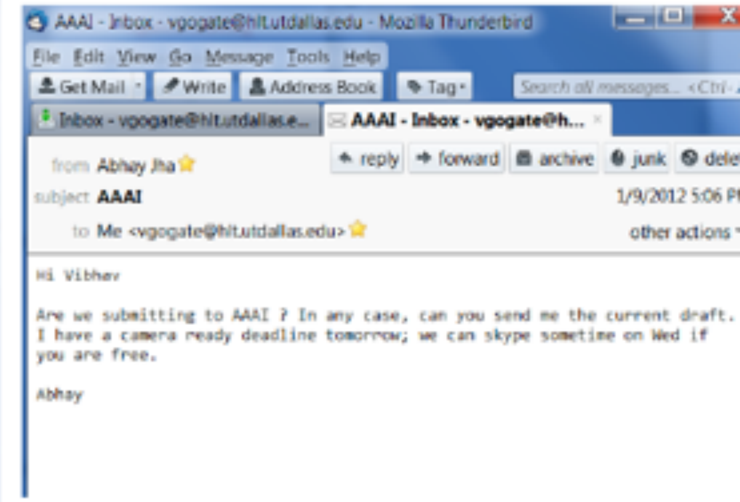
***Instructor: Prof. Stanislav Sobolevsky***
***Course Assistants: Tushar Ahuja, Maxim Temnogorod***


4
2
3
4
9
0
5
7
1
9
0
3
6
7
4

# Classification - examples

- Customer profiling
- Risk management
- Land use classification
- Digit recognition
- Image/face recognition
- Spam detection
- Weather patterns





Company home page

vs

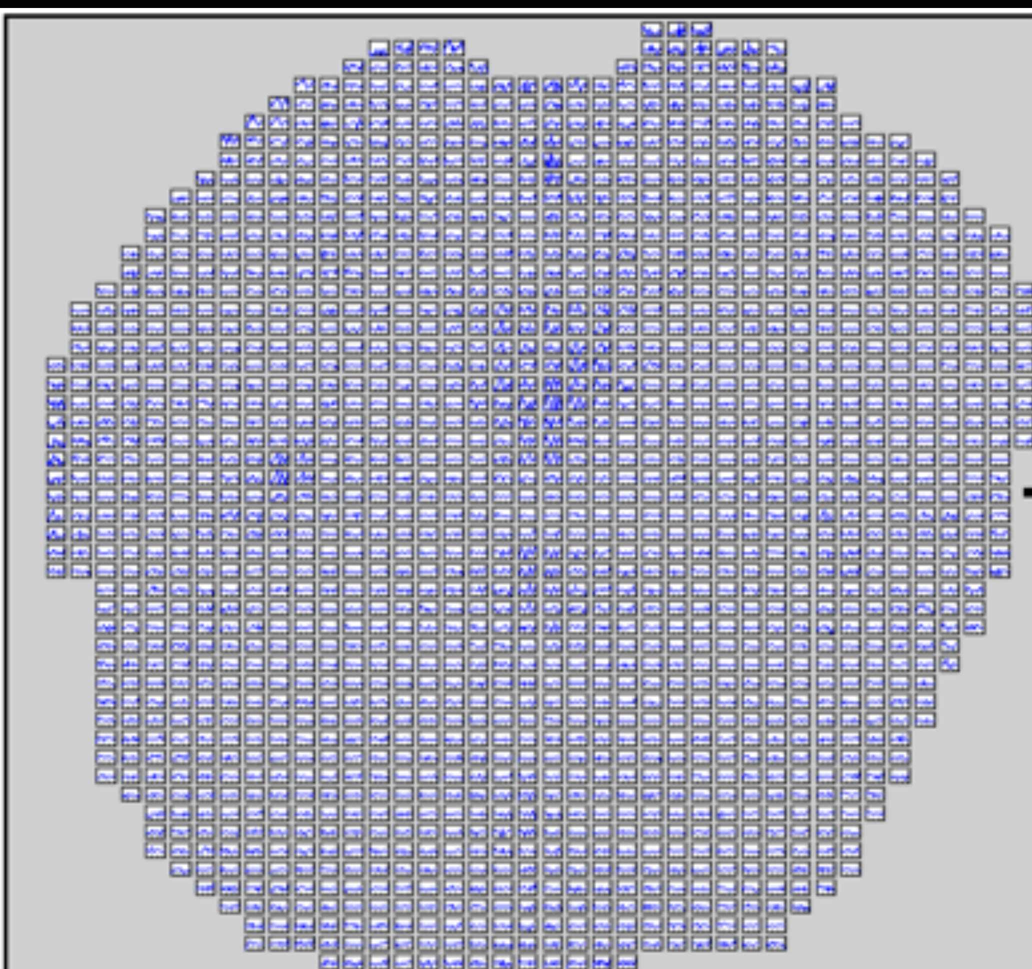
Personal home page

vs

University home page

vs

...



Reading a noun (vs verb)

[Rustandi et al., 2005]

# Supervised learning: classification

| Data/input | Discrete labels | → Dependence |
|---|---|---|
| $x_1$ | $y_1$ | $y = f(x)$ |
| | | $x^* \rightarrow y^*$ |
| $x_2$ | $y_2$ | y-discrete |
| $\ldots$ | $\ldots$ | $P(y\|x^*)$ |
| $x_N$ | $y_N$ | |

$$X = \{x_i, i = 1..N\} = \{x_i^j, i = 1..N, j = 1..n\}$$

$$Y = \{y_i, i = 1..N\}$$

# Binary/multiple classification

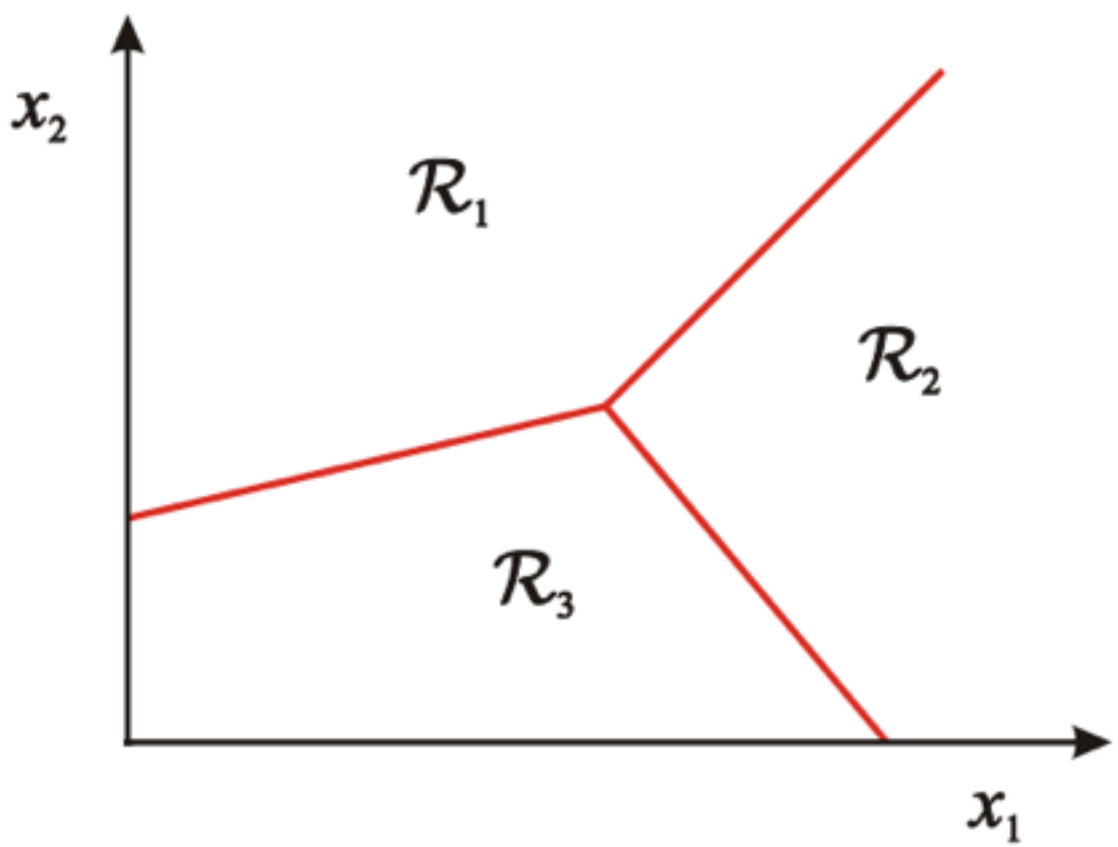
$$P(y|x^*) = Bern(y|\mu(x^*))$$

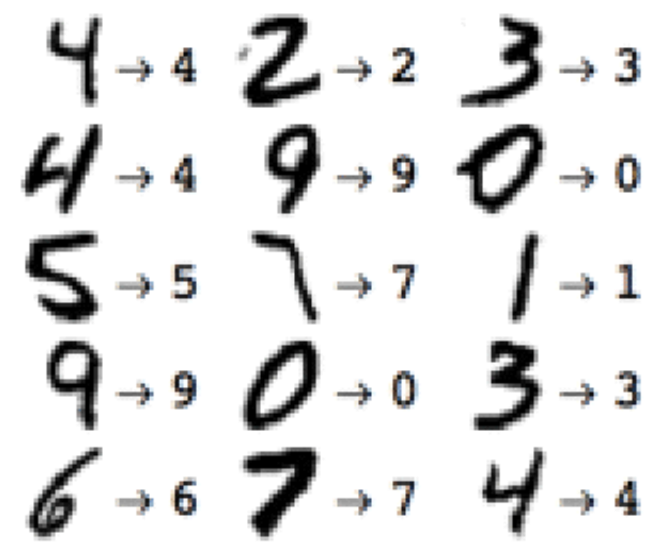
$$y = \begin{cases} 1 & \text{event happened} \\ 0 & \text{event not happened} \end{cases}$$

Multiple classification y=k

Multiple classification to binary:
c=k
c~=k

# Multi-class classification

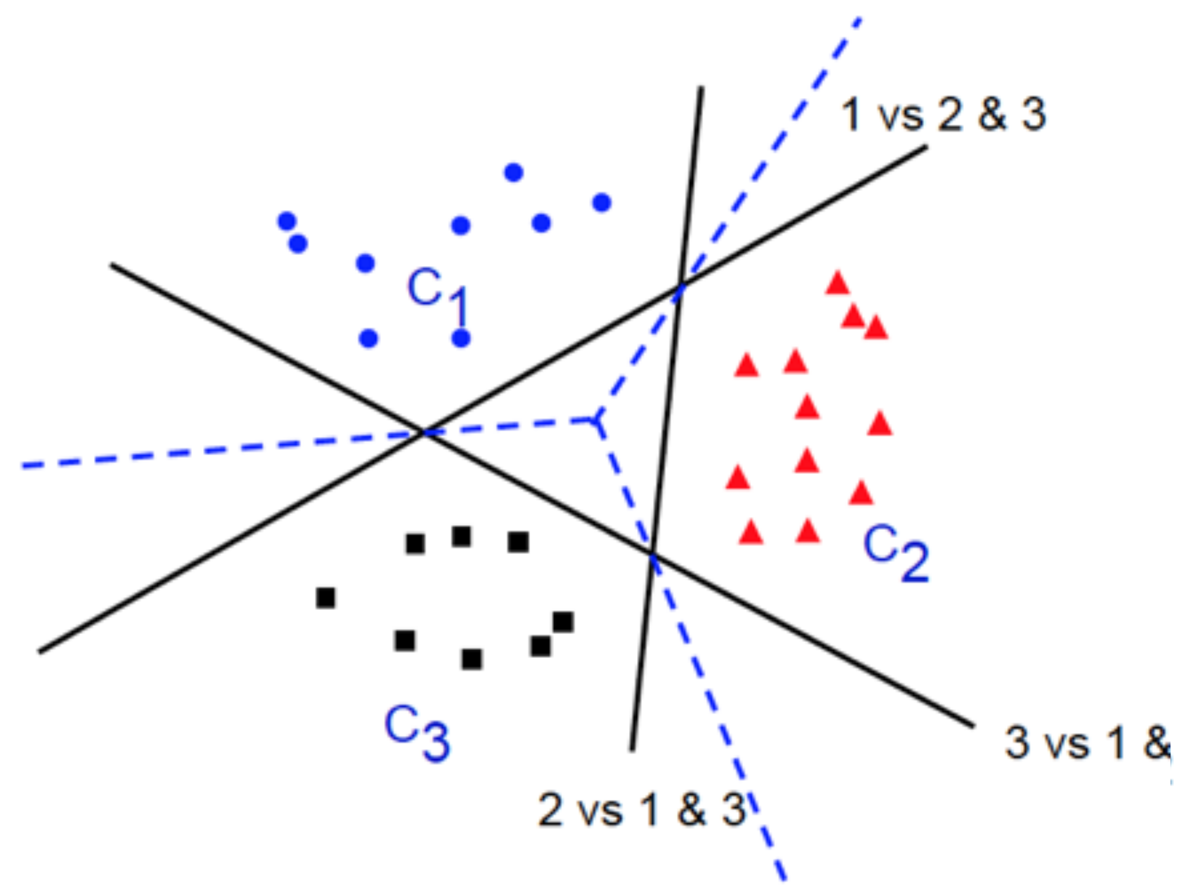


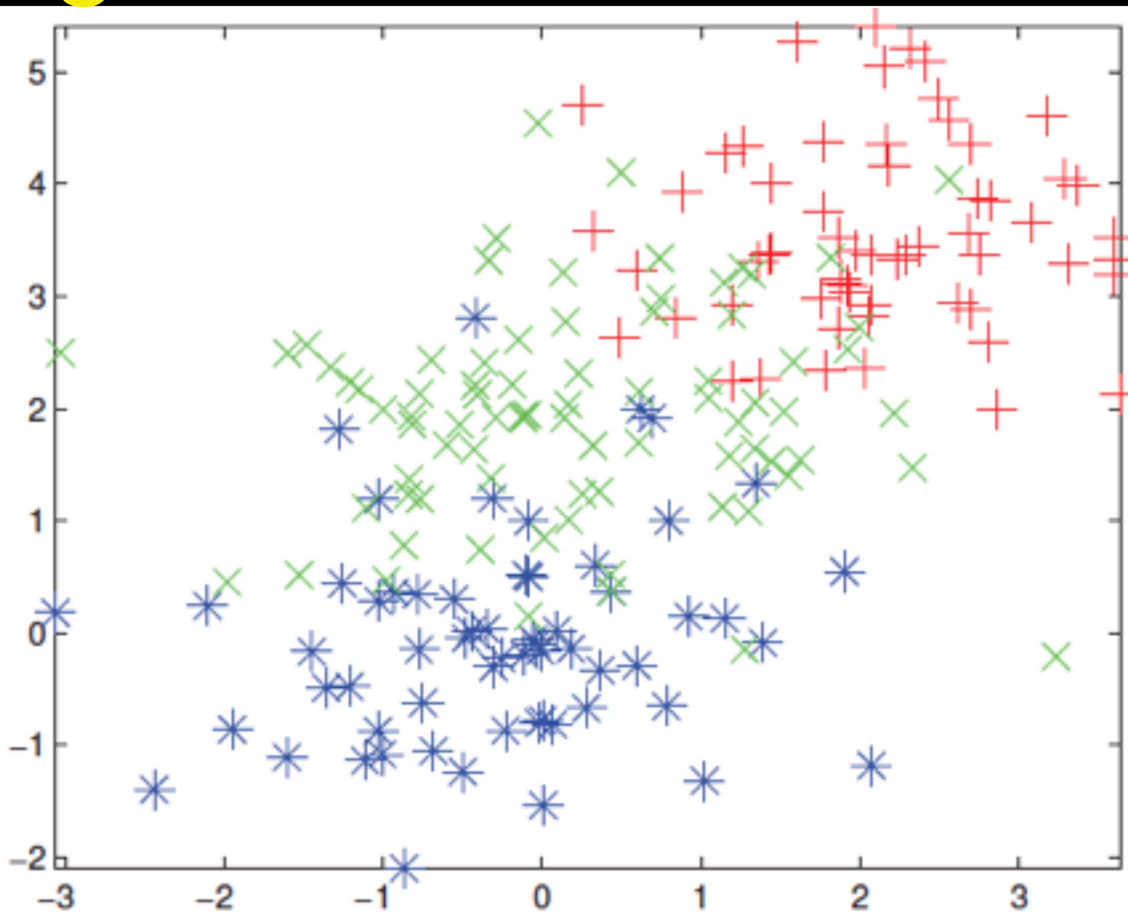

Is R1? Rather than R2 or R3

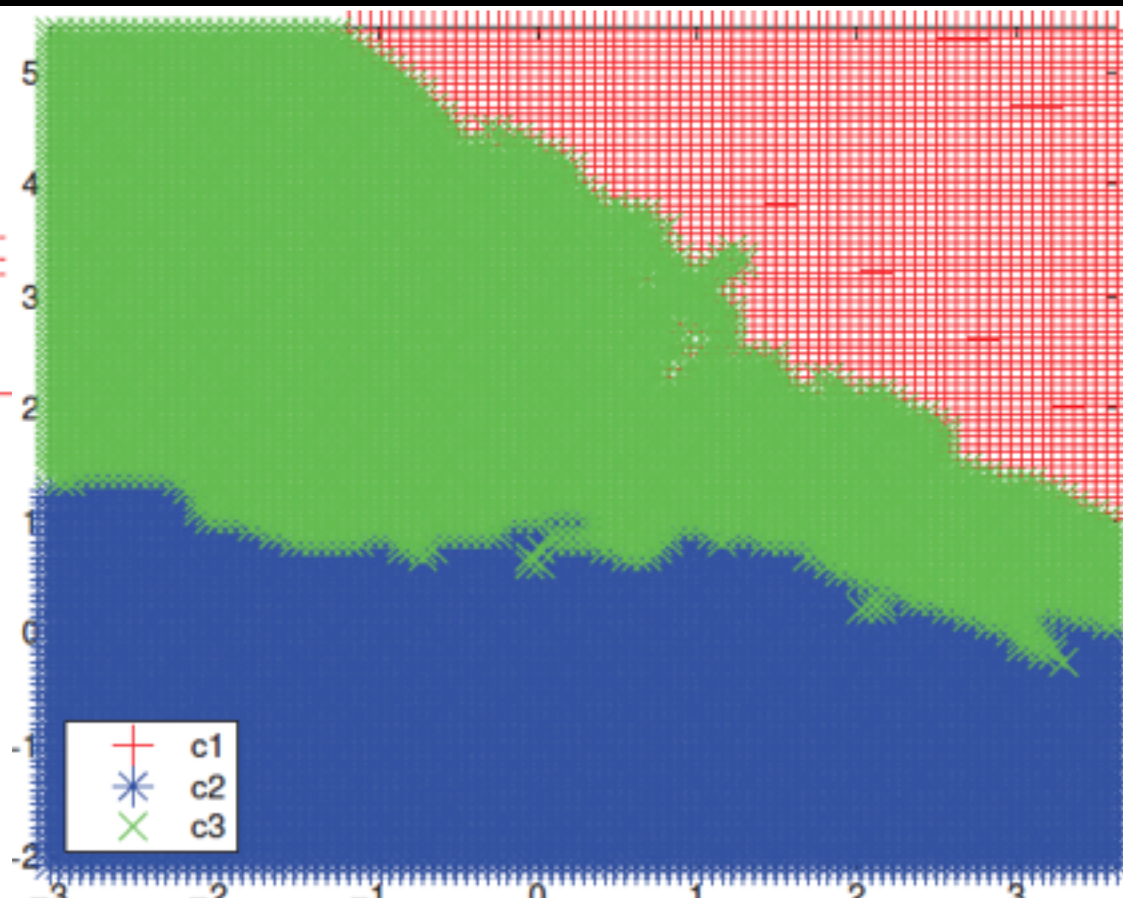
Is R2? Rather than R1 or R3

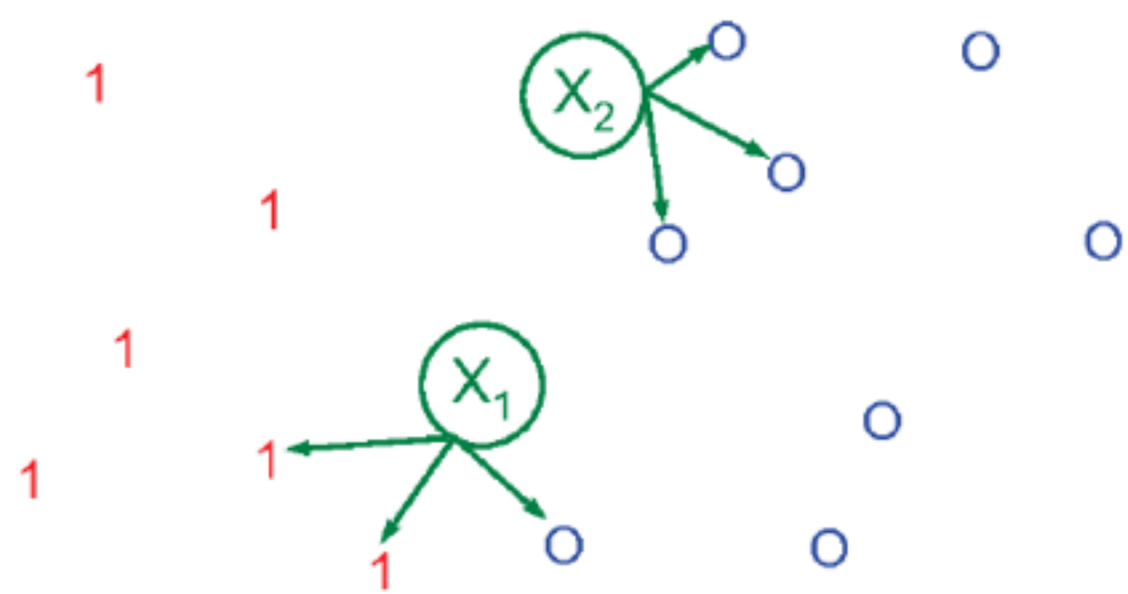
Is R3? Rather than R1 or R2

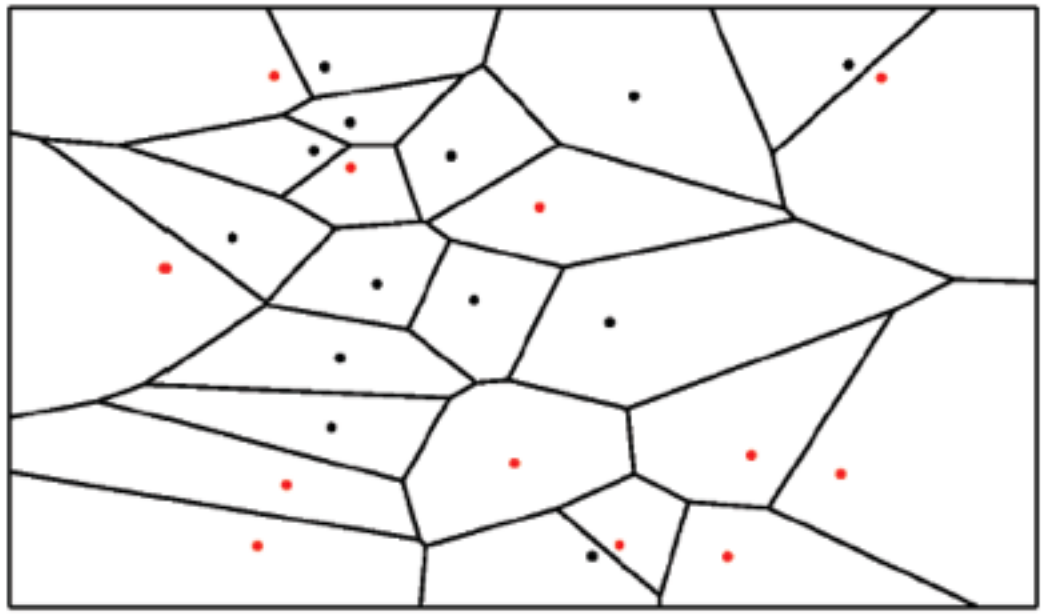


# K-neighbor classifier



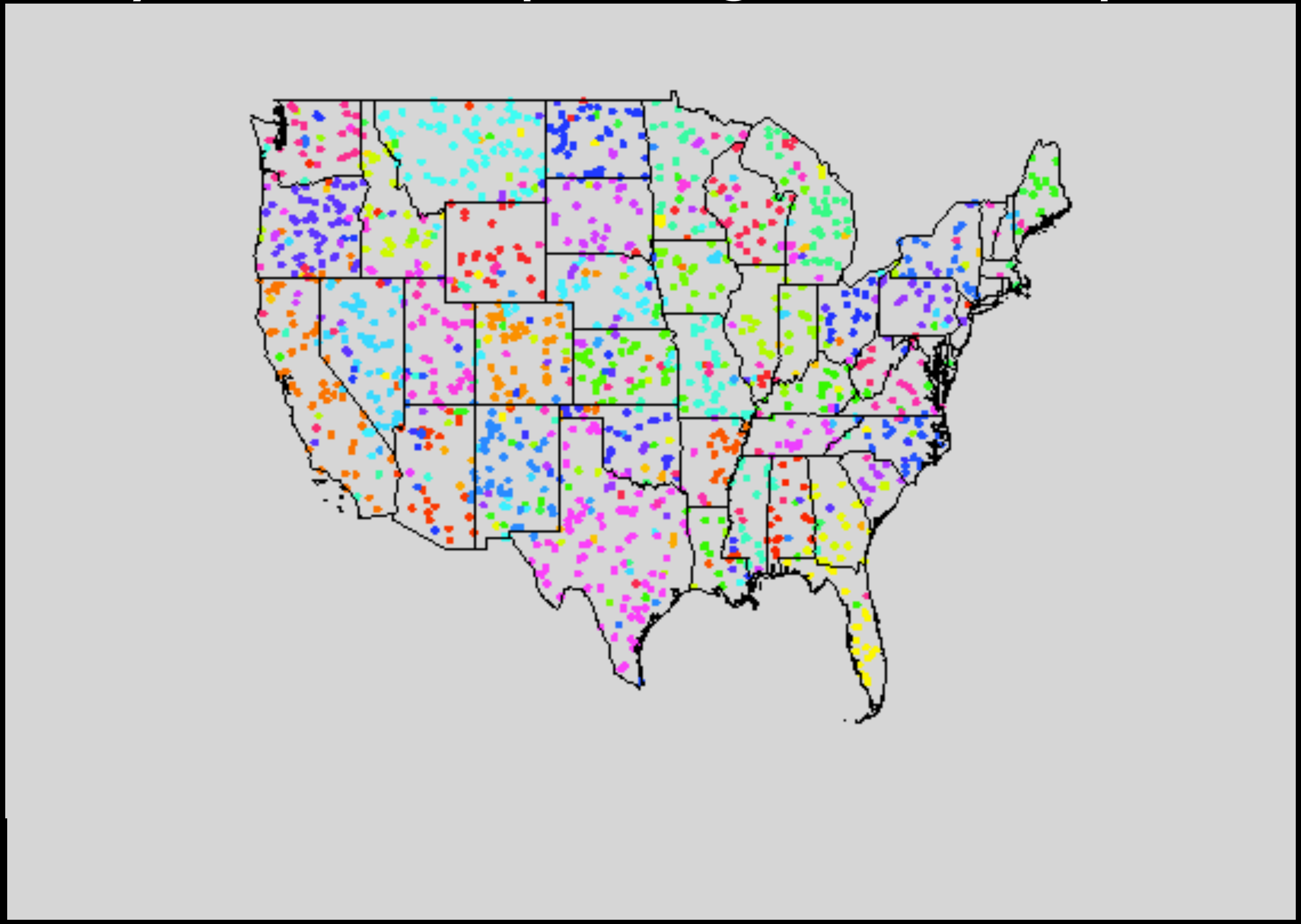



Murphy, Kevin P. *Machine learning: a probabilistic perspective*. MIT press, 2012

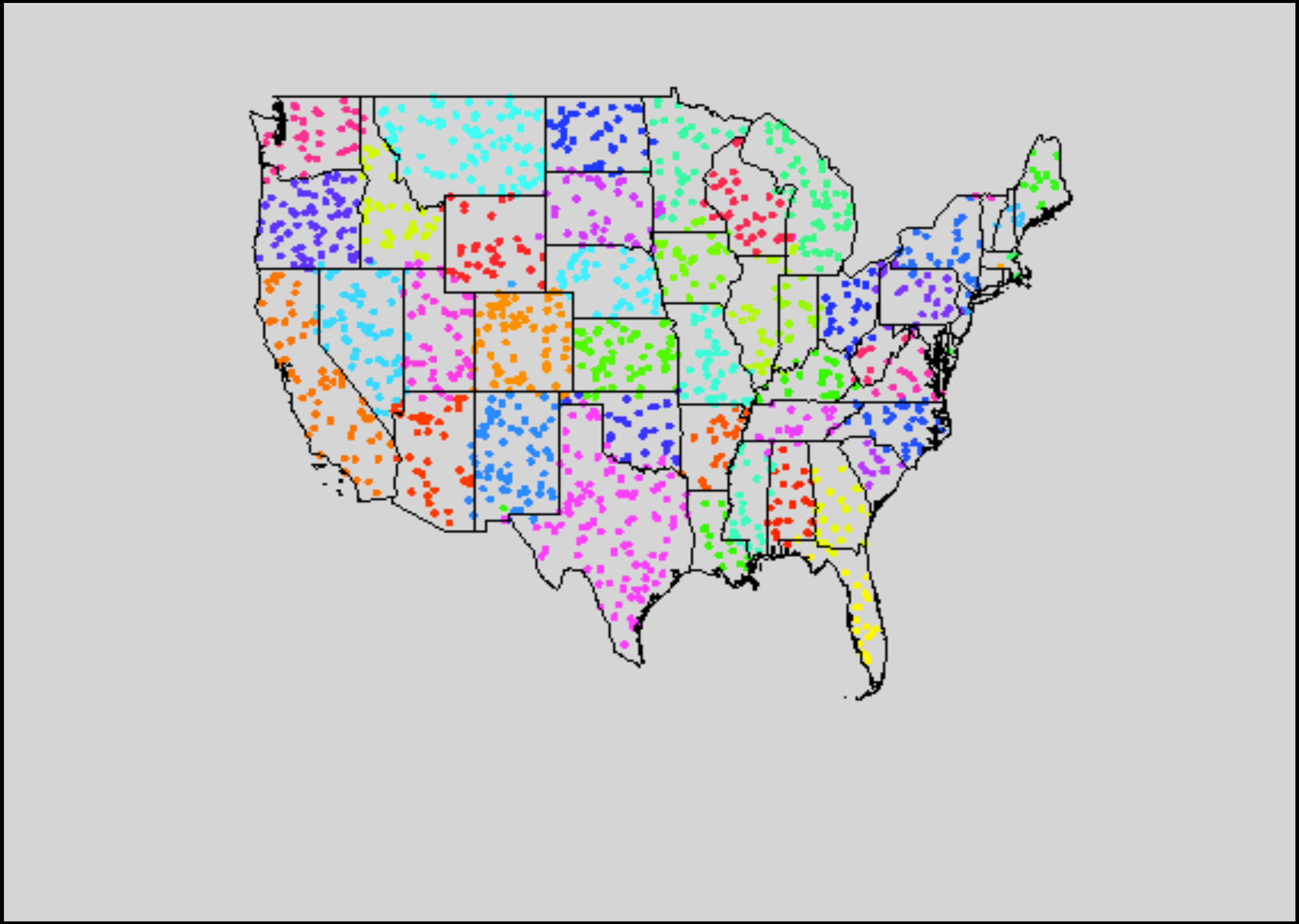
# 1.1.3. Example: K-neighbor classifier

What state you are in depending on the car plates around?

K=1



# 1.1.3. Example: K-neighbor classifier



K=10

# Spam classification "toy" example

| # | Cruise | Lottery | Win | Congratulations | Spam |
|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 0 | 1 | 1 | 1 | 1 |
| 3 | 1 | 1 | 0 | 0 | 1 |
| 4 | 1 | 0 | 1 | 1 | 1 |
| 5 | 1 | 0 | 0 | 0 | 0 |
| 6 | 0 | 0 | 0 | 1 | 0 |

| # | Cruise | Lottery | Win | Congratulations |
|---|---|---|---|---|
| 1 | 1 | 1 | 0 | 1 |
| 2 | 1 | 0 | 0 | 1 |

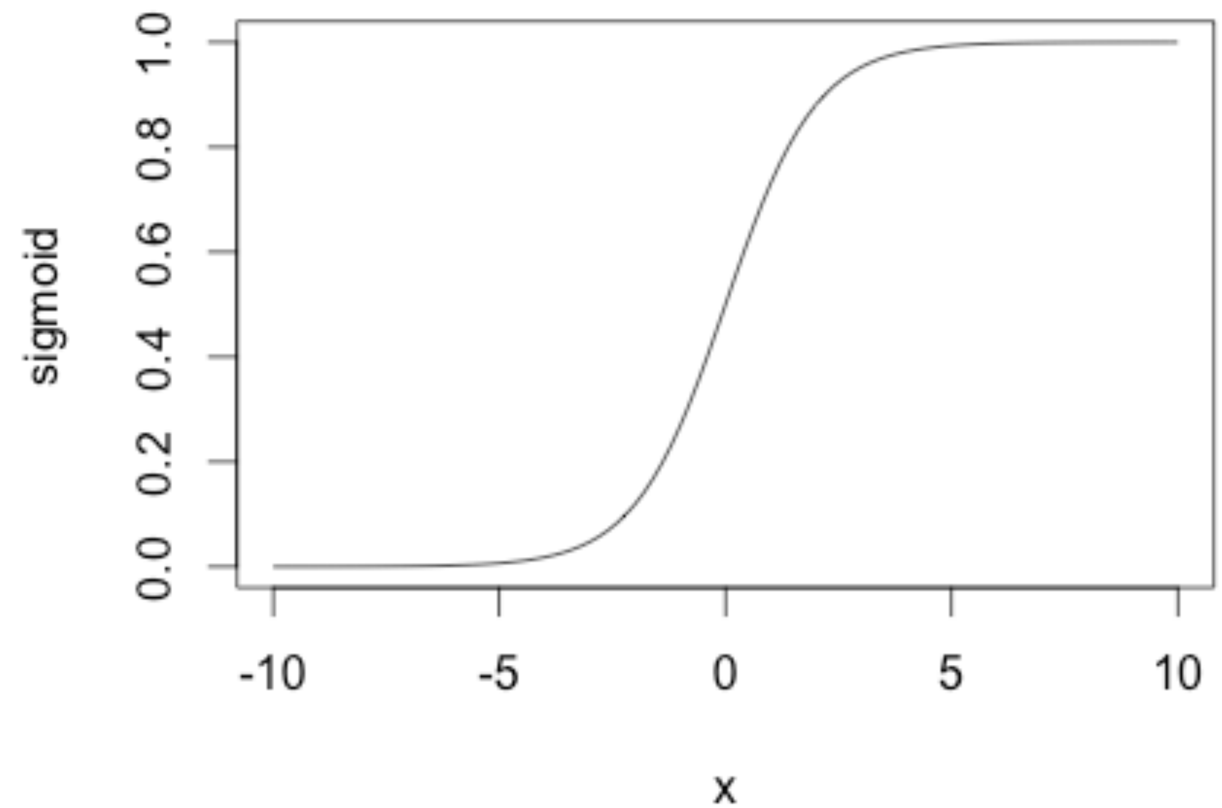
Spam or ham?

# Logistic regression

$$P(y|x,\beta) = Bern(y|\mu(x,\beta))$$

$$\mu(x,\beta) = f(x\beta)$$

$$f(x) = \sigma(x) = \frac{e^x}{1+e^x} = \frac{1}{1+e^{-x}}$$



# Logistic regression

$$P(y|x,\beta) = Bern(y|\mu(x,\beta))$$

$$f(x) = \sigma(x) = \frac{e^x}{1+e^x} = \frac{1}{1+e^{-x}}$$

$$\mu(x,\beta) = f(x\beta)$$

$$P(y=1) = \sigma(x\beta) = \frac{\exp(x\beta)}{1+\exp(x\beta)} = \frac{1}{1+\exp(-x\beta)}$$

$$P(y=0) = 1 - P(y=1) = \frac{1}{1+\exp(x\beta)}$$

# Logistic regression - log-likelihood

$$P(y=1)=\sigma(x\beta)=\frac{\exp(x\beta)}{1+\exp(x\beta)}=\frac{1}{1+\exp(-x\beta)}$$

$$L=\prod_i P(y=y_i|x_i,\beta)$$

$$P(y=0)=1-P(y=1)=\frac{1}{1+\exp(x\beta)}$$

$$log(L)=\sum_i log\left(P(y=y_i|x_i,\beta)\right)$$

$$=\sum_i y_i log\left(P(y=1|x_i,\beta)\right)+\sum_i(1-y_i)log\left(P(y=0|x_i,\beta)\right)$$

$$=-\sum_i log\left(1+exp((2y_i-1)x_i\beta)\right)$$

$$\beta=argmin_\beta\sum_i log\left(1+exp((2y_i-1)x_i\beta)\right)$$

# Accuracy

$$acc = \frac{|\{i : y_i^{est} = y_i^{true}\}|}{|\{i\}|}$$

Not all errors are the same!

Missing spam vs important e-mail as spam

False fire alarm vs missing a fire

# Types of outcomes

$$y = \begin{cases} 1 & \text{event happened} \\ 0 & \text{event not happened} \end{cases}$$

True

True positive: $y_i^{est} = y_i^{true} = 1$

True negative: $y_i^{est} = y_i^{true} = 0$

Errors

False positive: $y_i^{est} = 1, y_i^{true} = 0$

False negative: $y_i^{est} = 0, y_i^{true} = 1$

# Confusion matrix

$$\left[\begin{array}{c|c} TP & FN \\ \hline FP & TN \end{array}\right]$$

precision $$PPV = \frac{TP}{TP + FP}$$

true fires among reported fires

sensitivity or recall $$TPR = \frac{TP}{TP + FN}$$

reported fires among all fires

accuracy $$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$

fraction of true classifications