Right answers

Train

Predicted result

Test

Prediction

Label

Class

Features

Attributes

CISC 873 Class 2
8 class

15% is 60cm

- Confusion matrix

- Prediction accuracy ≠ 85% is not good

85% success in class A, 15% in class B

Less 85%

- Prediction accuracy ≠ right

Performance measures.
accuracy = \frac{(TP + TN)}{m}
\[ \text{Recall} = \frac{TP}{TP + FN} \]

\[ \text{Precision} = \frac{TP}{TP + FP} \]

F-Score = harmonic mean of precision
Cross-validation.

Area under curve.
compare accuracies - average these accuracies
Vanessa - effects of very busy model.

Simple

Bias in the model - too

Bias in the data

If the system

Bias - systematic misrepresentations

Data - how complex a model?
OVERFITTING

- model is too powerful
- learns exactly the training set
- model is too powerful
- learns exactly the training set

not that data
Clustering/Exploreation

By understanding

Similar

Why? But this is

have a difference will prediction be?
Clusters not necessarily true

I chose the similarity depend on clusters
$r_1 = (a_1, a_2)$

\[ \text{dist}(r_1, r_2) = \sqrt{(a_1 - a_2)^2 + (a_2 - a_2)^2} \]

& these are odd

\(m = 100\)

& high-dimensional spaces
prob that a reads as data?

\[
\left( \frac{1}{2} \right)^m
\]

The equation.

Prob that a record is close to uniform distributed.

Suppose my 0,1 in dataset.
Except hereditary disken

except

be let in many clusters to

Almost as closely arranged round it

$$\left( \frac{1}{2} \right)^w$$
process in science as he conceives.

Failing (a bit flakiness)
- not enough attributes?

- too many records? sample yourself?

- not enough records? sample Community Sample

...which records, which attributes?

data collection and cleaning

EXPE Ls

DONN
attributes

Complex - refers to a set of features

- repeat & average
- select value for each record
- custom attributes

Selection

- too many attributes? Attributes
Fixing missing values
Removing or adding attributes
Cleaning - replacing missing values

\[ a \cdot b \]

Correlation is not causation
- prediction (decision tree)
- clustering (K-means + anything)

2. Simple algorithms (play)

BE CONSERVATIVE
- which & why attribute selection
- patterns to preferential problems
- understanding

1. Statistics before anything
which needs an careful way.

- Perform evaluation

| M | M_2 | M_3 | M_4 |

- Prediction

- Perform feature selection

- Perform cross-validation (conservatively)

- Choose best match step

- Perform clusters
(MIME)

M, r, p them, (Whoon)

deployment (in the real world)