

Lecture title

Subject
Week

Me

melbournebioinformatics.org.au

Lecture title

1. Contents slide starts with a single #
2. Example algorithms

Normal slides start with a double ##

🍌 content goes here

Columns

left col

right col

Overprint

Clipped images

Code blocks

```
put anything in the  
class name and you  
get grey
```

```
1 add .numberLines for  
2 numbered lines
```

```
1 work out bash/python
```

👑 it should match the text
indent

Alert blocks

Alerted block

- 👤 alerted content
- 👤 use the three colons to break out of the alertblock

We also have unstyled H3 blocks

- 👤 this matches the layout of alert blocks
- 👤 but has no colour

You need to use latex code to style the headers

- 👤 use the contrast sparingly for highlights

Lecture title

1. Contents slide starts with a single #
2. Example algorithms

Example algorithms

K-means (D, k, ε)

```
1  $t = 0$ 
2 Randomly initialize  $k$  centroids:  $\mu_1^t, \mu_2^t, \dots, \mu_k^t \in \mathbb{R}^d$ 
3 repeat
4    $t \leftarrow t + 1$ 
5    $C_j \leftarrow \emptyset$  for all  $j = 1, \dots, k$ 
6   // Cluster assignment step
7   foreach  $x_j \in D$  do
8      $j^* \leftarrow \arg \min_i \left\{ \|x_j - \mu_i^t\|^2 \right\}$  // Assign  $x_j$ 
9     to closest centroid
10     $C_{j^*} \leftarrow C_{j^*} \cup \{x_j\}$ 
11    // Centroid update step
12    foreach  $i = 1$  to  $k$  do
13       $\mu_i^t \leftarrow \frac{1}{|C_i|} \sum_{x_j \in C_i} x_j$ 
14  until  $\sum_{i=1}^k \|\mu_i^t - \mu_i^{t-1}\|^2 \leq \varepsilon$ 
```

Procedure Dbscan($X, \varepsilon, \text{minpts}$)

```
1 foreach unvisited point  $x \in X$  do
2   mark  $x$  as visited
3    $N \leftarrow \text{GetNeighbours}(x, \varepsilon)$ 
4   if  $|N| < \text{minpts}$  then
5     mark  $x$  as noise
6   else
7      $C \leftarrow \{x\}$ 
8     foreach point  $x' \in N$  do
9        $N \leftarrow N \setminus x'$ 
10      if  $x'$  is not visited then
11        mark  $x'$  as visited
12         $N' \leftarrow \text{GetNeighbours}(x', \varepsilon)$ 
13        if  $|N'| \geq \text{minpts}$  then
14           $N \leftarrow N \cup N'$ 
15        if  $x'$  is not yet member of any cluster
16          then
17             $C \leftarrow C \cup \{x'\}$ 
```

label equations

Non-negativity: $d(a, b) \geq 0$

Identity: $d(a, a) = 0$

Symmetry: $d(a, b) = d(b, a)$

Triangle inequality: $d(a, c) \leq d(a, b) + d(b, c)$

Conditional probability:

$$P(A | B) = \frac{P(A \cap B)}{P(B)}$$

Probability of A and B

Probability of A given B

Probability of B

tables

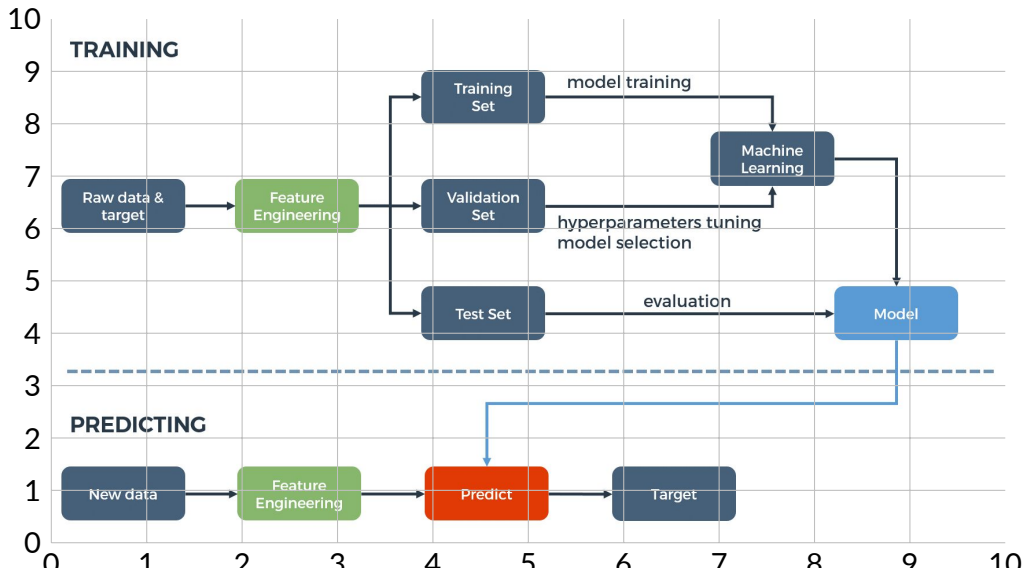
It's safe to use markdown in the table cells, e.g. [like this](#).

Sequences	Unrooted trees
3	1
4	3
5	15
10	>
	2 000 000

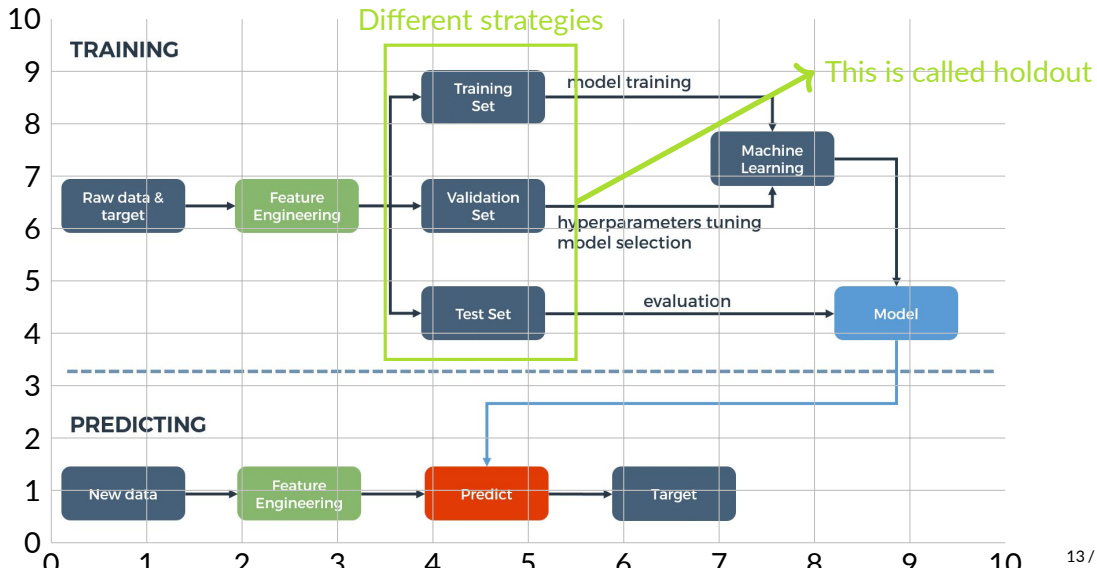
Suppress column headers

S1	ACTGTG
S2	TCACAG
S3	AGTCAG
S4	AGTGTC
S5	TCAGTG

Annotated figure



Annotated figure



Nonlinear dimensionality reduction

The itemize length in the right column is set to the length of Geodesic: + \labelsep as follows:

```
\newlength{\somelength}  
\settowidth{\somelength}{Geodesic:}  
\setbeamerwidth{description width=\somelength}  
\setlength{\leftmargini}{\somelength + \labelsep}
```

This lines the bullets up with the description labels... nice!

Manifold: a nonlinear low-dimensional surface

- data often lies on or near manifolds

ISOMAP (Isometric Feature Mapping):

- preserves the global, non-linear geometry of the data by preserving the **geodesic** distances

Geodesic: shortest route between two points on the surface of the manifold