Computational Genomics (0382.3102)
Lecture 6

Brief Intro to Machine Learning, and to Suffix Trees

Prof. Benny Chor
School of Computer Science
Tel-Aviv University

Lecture based in part on chapter 1 in Scholkopf & Smola’s book, Learning with Kernels, and on a ppt presentation by Prof. Haim Kaplan (Tel-Aviv Univ.)
First - Dry Run of Approx. MSA

- Four Sequences: \( S_1, S_2, S_3, S_4 \)
  \[ d(\text{mismatch}) = 3, \quad d(\text{indel}) = 2. \]
First - Dry Run of Approx. MSA

- Four Sequences: $S_1, S_2, S_3, S_4$
  $d(\text{mismatch}) = 3$, $d(\text{indel}) = 2$. 
First - Dry Run of Approx. MSA

- Four Sequences: $S_1, S_2, S_3, S_4$
  $d(\text{mismatch}) = 3$, $d(\text{indel}) = 2$.

- $S_1 = \text{AAAAAAAAAAAAAAAA}$
  $S_2 = \text{AAAAAAAAAAAAAC}$
  $S_3 = \text{AAAAAAAAAGAAAAA}$
  $S_4 = \text{TAAAAAAAAAAAA}$
First - Dry Run of Approx. MSA

• Four Sequences: $S_1, S_2, S_3, S_4$
  $$d(\text{mismatch}) = 3, \ d(\text{indel}) = 2.$$  

• $S_1 = \text{AAAAAAAAAAAAAAAA}$
  $S_2 = \text{AAAAAAAAAAAAAC}$
  $S_3 = \text{AAAAAAAAAGAAAAA}$
  $S_4 = \text{TAAAAAAAAAAAAAA}$

• $d(S_1, S_i) = 2, \ d(S_i, S_j) = 6.$
First - Dry Run of Approx. MSA

- Four Sequences: $S_1, S_2, S_3, S_4$
  \[ d(\text{mismatch}) = 3, \ d(\text{indel}) = 2. \]
- $S_1 = \text{AAAAAAAAAAAAAA}
  S_2 = \text{AAAAAAAAAAAAAC}
  S_3 = \text{AAAAAAAAAGAAAAA}
  S_4 = \text{TAAAAAAAAAAAA}

- $d(S_1, S_i) = 2, \ d(S_i, S_j) = 6.$
- So $S_1$ is the “center”.

Alg. Progressive Alignments
Alg. Progressive Alignments

1) \[ S_1 = \text{AAAAAA} \cdots \text{AAAA} - \]
   \[ S_2 = \text{AAAAA} \cdots \text{AAAAAA} \text{C} \]
Alg. Progressive Alignments

1) \[ S_1 = \text{AAAAAAAAAAAAA} - \]
   \[ S_2 = \text{AAAAAAAAAAAAAC} \]
Alg. Progressive Alignments

1) \( S_1 = \text{AAAAAAAAAAAAA} - \)
   \( S_2 = \text{AAAAAAAAAAAAA}C \)

2) \( S_1 = \text{AAAAAA} - \text{AAAAA} - \)
   \( S_2 = \text{AAAAAA} - \text{AAAAAC} \)
   \( S_3 = \text{AAAAAAAGAAAA} - \)
Alg. Progressive Alignments

1) \( S_1 = \text{AAAAAA} \cdots \text{AAAAA} \cdots \text{-} \)
   \( S_2 = \text{AAAAAA} \cdots \text{AAAAA} \cdots \text{C} \)

2) \( S_1 = \text{AAAAAA} \cdots \text{AAAAA} \cdots \text{-} \)
   \( S_2 = \text{AAAAAA} \cdots \text{AAAAA} \cdots \text{C} \)
   \( S_3 = \text{AAAAAA} \text{G} \cdots \text{AAAAA} \cdots \text{-} \)
Alg. Progressive Alignments

1) \( S_1 = \text{AAAAAAAAAAAAAA} - \)
\( S_2 = \text{AAAAAAAAAAAAAC} \)

2) \( S_1 = \text{AAAAAAA} - \text{AAAAAA} - \)
\( S_2 = \text{AAAAAAA} - \text{AAAAAAC} \)
\( S_3 = \text{AAAAAAAAGAAAAAA} - \)

3) \( S_1 = \text{AAAAAAA} - \text{AAAAAA} - \)
\( S_2 = \text{AAAAAAA} - \text{AAAAAAC} \)
\( S_3 = \text{AAAAAAAAGAAAAAA} - \)
\( S_4 = \text{TAAAAAAA} - \text{AAAAAA} - \)
Alg. Progressive Alignments

1) $S_1 = \text{AAAAAAAAAAAA} -$  
   $S_2 = \text{AAAAAAAAAAAAAC}$

2) $S_1 = \text{AAAAAAA} -$ $\text{AAAAAA} -$  
   $S_2 = \text{AAAAAAA} -$ $\text{AAAAAAC}$  
   $S_3 = \text{AAAAAAAGAAAAA} -$ 

3) $S_1 = \text{AAAAAAA} -$ $\text{AAAAAA} -$  
   $S_2 = \text{AAAAAAA} -$ $\text{AAAAAAC}$  
   $S_3 = \text{AAAAAAAGAAAAA} -$  
   $S_4 = \text{TAAAAAAA} -$ $\text{AAAAAA} -$
Pluto - General Background

Pluto, our loyal shepherd dog,
is highly intelligent and alert.

But where is Pluto?
Learning – Basic Scenario

• Pluto keeps an eye on a herd of sheep.
Learning – Basic Scenario

- Pluto keeps an eye on a herd of *sheep*.
- There are white sheep and black sheep.
Learning – Basic Scenario

• Pluto keeps an eye on a herd of sheep.
• There are white sheep and black sheep.
• New sheep arrive at night time, carried by storks.
Learning – Basic Scenario

- Pluto keeps an eye on a herd of *sheep*.
- There are *white* sheep and *black* sheep.
- New sheep arrive at night time, carried by *storks*.
- Pluto should be able to identify the newly arriving sheep and report their color to the herd’s *CEO*. 

© Benny Chor
Learning – Basic Scenario

- Pluto keeps an eye on a herd of *sheep*.
- There are **white** sheep and **black** sheep.
- New sheep arrive at night time, carried by **storks**.
- Pluto should be able to identify the newly arriving sheep and report their color to the herd’s **CEO**.
- Nights are pitch dark.
Learning – Basic Scenario

- Pluto keeps an eye on a herd of *sheep*.
- There are *white* sheep and *black* sheep.
- New sheep arrive at night time, carried by *storks*.
- Pluto should be able to identify the newly arriving sheep and report their color to the herd’s *CEO*.
- Nights are pitch dark.
- Even though Pluto is equipped with the latest night vision technology, it cannot see a thing . . . ahmmmm . . . sheep.
Learning – Basic Scenario

- Pluto keeps an eye on a herd of *sheep*.
- There are *white* sheep and *black* sheep.
- New sheep arrive at night time, carried by *storks*.
- Pluto should be able to identify the newly arriving sheep and report their color to the herd’s *CEO*.
- Nights are pitch dark.
- Even though Pluto is equipped with the latest night vision technology, it cannot see a thing . . . ahmmm . . . sheep.
- Fortunately, newly arriving sheep are *quite noisy*, and Pluto’s has great senses of *hearing* and *directionality*.
The Sheep(ing) News

- Sheep of different colors do not mingle together.
The Sheep(ing) News

- Sheep of different colors do not mingle together.
- The same applies to the newcomers.
The Sheep(ing) News

- Sheep of different colors do not mingle together.
- The same applies to the newcomers.
- Pluto learns all sheep’s facts (e.g. location and colors) at daytime.
The Sheep(ing) News

- Sheep of different colors do not mingle together.
- The same applies to the newcomers.
- Pluto learns all sheep’s facts (e.g. location and colors) at daytime.
- Pluto then applies this knowledge to infer new sheep’s colors from their location.
Pluto, a Learning Expert

...the story of the sheep dog who was herding his sheep, and serendipitously invented both large margin classification and Sheep Vectors...
A Crude Classifier
Better Classifier
Better (Too Good?) Classifier
Optimal Margins
Back to Pluto

- Pluto’s task would be rather easy if white and black sheep could be separated by a line in the 2D plane (in general, a hyperplane in $k$).
Back to Pluto

- Pluto’s task would be rather easy if white and black sheep could be separated by a line in the 2D plane (in general, a hyperplane in $k$).
- Unfortunately, life is not always so linear
Back to Pluto

- Pluto’s task would be rather easy if white and black sheep could be separated by a line in the 2D plane (in general, a hyperplane in $k$).
- Unfortunately, life is not always so linear
Solution: Transform!