

Homework

Construct an HMM to identify a pattern in DNA sequences

Given: 100 “true” training sequences
100 “false” training sequences

Your HMM will be applied to 1,000 testing sequences, some fraction of which contain the pattern.

The pattern: 3-5 11bp segments
strong, but degenerate sequence
first segment not more than 100bp from start
spacing between segments is 10-30bp

Homework

You must turn in:

1. A program which reads the file of testing sequences and writes two files:

res-true.txt - names of sequences your program considers “true”

res-false.txt - names of sequences your program considers “false”

Files should include your HMM score

```
> testing 1 0.957
> testing 2 0.831
> testing 3 0.902
```

2. A schematic (state diagram) of your HMM architecture

PDF preferred