Namn och Personnummer

Instruktioner

Tillåtna hjälpmedel: papper och penna.

Varje fråga skall besvaras inom det befintliga utrymmet, om ni behöver mer utrymme skriv på baksidan.

Glöm inte att skriva namn och personnummer på alla papper. Om dessa saknas kan det hända att dina svar inte tas med i din totalpoäng!!

Du kan svara på alla frågor på svenska eller engelska.

För Betyg 3 krävs 15 poäng, Betyg 4 20 poäng och Betyg 5 25 poäng.

Instructions

Allowed tools: Paper and Pen and Dictionary.

Each question should be answered within the provided space, if you need more space write on the back of the paper.

Do not forget to write name and personal number on each page, if this information is missing the answers may not count !!

You may answer in English or Swedish

For Grade 3 15p is needed, Grade 4 20p and Grade 5 25p.

Namn och Personnummer

1. Bilden visar HMM-arkitekturen för en prediktor (TMHMM) för membranproteintopologi. De olika tillstånden (rutorna) karakteriseras av bl.a. olika aminosyrakompositioner. Vilka slags aminosyror (hydrofoba, hydrofila, aromatiska, positivt laddade och negativt laddade) är över- respektive underrepresenterade i de olika tillstånden (3p)?

Namn och Personnummer

2. När man skall träna ett neuralt nätverk av feed-forward-typ för att lära det känna igen ett proteinsekvens-motiv är det flera begrepp som är viktiga att känna till. Beskriv följande begrepp:

   (a) gles kodning ("sparse encoding") av sekvensdata (1p),
   (b) kors-validering ("cross-validation") (1p),
   (c) överträning (1p).

Namn och Personnummer

3. Beskriv generellt hur heuristiska sökmetoder för sekvenssökningar fungerar (t.ex. de som används i BLAST eller FASTA) med utgångspunkt från hur dynamisk programmering fungerar. (Tips tänk på hur man hittar identiska regioner). (1p) Beskriv för och nackdelarna med heuristiska metoder. (2p).

Namn och Personnummer

4. Förklara vilka för- och nackdelar en Hidden Markov Model (HMM) har jämfört med ett reguljärt uttryck (typ PROSITE) för att hitta mönster i en proteinsekvens. (3p)

Namn och Personnummer

5. När är det lämpligt att använda globala respektive lokala alignments ? Beskriv deras tillämpningar. (2p)

Namn och Personnummer

6. It has just been discovered that mutations in the gene coding for protein "P" are the primary cause of disease "Z". Sequence database searches show that "P" belongs to a family of uncharacterised proteins. A search of Pfam suggests that the first 150 residues of the protein are part of the conserved domain family "D" which is found in other multi-domain proteins. The molecular mechanism(s)/function(s) of "D" is not known. We will assume that the remaining 200 residues form domain "E". Using PSI-BLAST, no similarity to sequences of known three-dimensional structure can be found for any part of "P".

   - How could you use the Pfam database to suggest (guess) a possible (and very general) function for "P"? [1p]
   - Which structure prediction technique is most applicable to protein "P"? Assuming that this method gave you a reliable result, how can you use it to suggest a function for "P"? [2p]
   - What could/should you do to the amino-acid sequence of "P" before sending it to a structure prediction server on the web? [2p]

Namn och Personnummer

7. Problemet att identifiera gener utan att använda information från andra sekvenser i en genomisk DNA-sekvens är mycket svårare för eukaryota genom än för prokaryota. Ange två separata anledningar till varför det är så. (2p)

8. Med hjälp av databasen Pfam kan man göra statistik över förekomst (finns en viss familj?) och frekvens (hur många proteiner tillhör familjen?) av olika proteinfamiljer i ett genom. Om man jämför arter som t.ex. bananflugan (Drosophila), nematod-masken (C. elegans) och människa, varierar förekomst eller frekvens mest mellan arterna? Förklara kort, gärna med exempel. (2p)

Namn och Personnummer

9. Du har fått en nukleotidsekvens (och dess proteinsekvens) från ett experiment. Det enda du vet är att det är en gen från människa. Vilka två databaser skulle du välja att söka i först för att få en uppfattning om vad det kan vara för gen? Förklara kort varför. (2p)

Namn och Personnummer

10. Vad är “CASP” och varför har “CASP” varit viktigt för proteinstrukturprediktions området ? (2p)

Namn och Personnummer

11. Below you find one rooted and four unrooted phylogenies for two species of fish (shark and herring), one reptile (crocodile) and two mammals (kangaroo and human).

    i) If you remove the root from the Rooted Tree, which of the four Unrooted Trees does it correspond to? (1p)

    ii) Consider the Rooted Tree. Are the following groups monophyletic or not: a) Fish (i.e., shark and herring) b) mammals (i.e., kangaroo and human)? (2p)