

Programa del curso - Algoritmos en Bioinformática. Verano de 2011 (febrero-marzo)

Ámbito de aplicación y la forma:

El alcance del curso es dar a los estudiantes una introducción detallada a la bioinformática y los algoritmos de la bioinformática utilizados para la construcción de los métodos de predicción de alta calidad. El curso consta de conferencias, discusiones, ejercicios y proyectos. El curso es evaluado en términos de ejercicios y un informe escrito que será entregado en el final del semestre y un examen oral.

Duración del Curso:

2 semanas

Tipo de evaluación:

El examen oral y el informe escrito. Informe escrito (50%) y examen oral (50%).

Objetivos generales:

Para proporcionar al alumno una visión general y profunda comprensión de algoritmos de aprendizaje máquina de bioinformática. Permitir que el estudiante primero evaluar cual algoritmo será el más adecuado para responder a una cuestión biológica determinada y próxima aplicación y desarrollo de herramientas de predicción basadas en este tipo de algoritmo para describir los complejos problemas biológicos, como las reacciones del sistema inmunológico, el descubrimiento de la vacuna, la estructura de proteínas y función, etc..

Objetivos de aprendizaje:

Un estudiante que ha cumplido los objetivos del curso será capaz de:

- Entender los detalles de los algoritmos utilizados en la bioinformática.
- Desarrollar programas de ordenador la aplicación de estos algoritmos.
- Identificar el tipo de algoritmo más adecuado para describir un problema biológico determinado.
- Comprender los conceptos de redundancia de datos y la reducción de la homología.
- Desarrollar algoritmos de predicción de la bioinformática que describe un problema biológico determinado.
- Aplicación y desarrollo de herramientas de predicción en un nivel de detalle utilizando los siguientes algoritmos: Programación dinámica (Dynamic programming), la agrupación de secuencias (Sequence clustering), las matrices de peso (Weight matrices), redes neuronales artificiales (Artificial neural networks), modelos ocultos de Markov (Hidden Markov models),
- Diseño de un proyecto en el que un problema biológico se analiza utilizando uno o varios algoritmos de aprendizaje automático.
- Implementar, documentar y presentar el proyecto de curso.

Contenido:

El curso cubrirá los algoritmos más utilizados en la bioinformática inmunológica. Se hará hincapié en la aplicación matemática precisa de los algoritmos en términos de programas de ordenador funcional. Durante el curso, los problemas biológicos serán presentados y analizados con el propósito de poner de relieve los puntos fuertes y débiles de los diferentes algoritmos. Los siguientes temas serán cubiertos:

- Programación dinámica (Dynamic programming): Needleman-Wunsch, Smith-Waterman, and alignment heuristics
- La redundancia de datos y la reducción de la homología (Data redundancy and homology reduction): Hobohm and other clustering algorithms
- Matrices Peso (Weight matrices): Sequence weighting, pseudo count correction for low counts, Gibbs sampling, and Psi-Blast
- Modelos ocultos de Markov (Hidden Markov Models): Model construction, Viterbi decoding, and posterior decoding, and Baum Welsh HMM learning
- Las redes neuronales artificiales (Artificial neural networks): Architectures and sequence encoding, feed-forward algorithm, and back propagation

El curso constará de conferencias, sesiones de discusión y ejercicios en el computador, donde los estudiantes serán introducidos a los diferentes algoritmos, su aplicación y uso en el análisis de los problemas biológicos.

Responsable:

Morten Nielsen, mniel@cbs.dtu.dk

Program 2011

Day 1: Course introduction

Lectures

- Introduction to course
- Immune system overview
- UNIX crash course 101
- C-programming crash course 101

Exercise

- C programming using pointers and linked lists

Day 2: Weight matrix construction

Lectures

- Weight matrix (PSSM) construction,
- Gibbs sampling, and
- Psi-Blast

Exercise

- Implementation of PSSM construction from pre-aligned sequences including pseudo count correction for low counts and sequence clustering

Day 3: Sequence alignment

Lectures

- Blosom scoring matrices
- Sequence alignment
- Data redundancy reduction algorithms (Hobohm1 and Hobohm2).

Exercise

- Implementation of Smith-Waterman Dynamic programming algorithm
- Implementation of data homology reduction algorithms

Day 4: Hidden Markov Models

Lectures

- Hidden Markov models introduction
- Viterbi decoding
- Forward/Backward algorithm, Posterior decoding, Baum-Welsh learning
- Profile Hidden Markov Models.

Exercise

- Implementation of Viterbi and posterior decoding
- Training of a profile Hidden Markov model using HMMer

Day 5: Artificial neural networks

Lectures

- Background
- Sequence encoding
- Feed forward algorithm
- Back-propagation and neural network training

- Network training - cross-validation

Exercise

- Implementation of sequence encoding feed forward and back-propagation algorithms

Day 6-9: Project work

Day 10: Exam