

What is a Project

- Purpose
 - Use a method introduced in the course to describe some biological problem
 - How
 - Construct a data set describing the problem
 - Define which method to use
 - (Develop method)
 - Train and evaluate method
 - (Compare performance to other methods)
 - Documentation
 - Make a poster describing project
 - Groups of 3-4 people
-

Project list

1. Peptide MHC binding predictions using position specific scoring matrices including pseudo counts and sequences weighting clustering (Hobohm) techniques
 2. Peptide MHC binding predictions using artificial neural networks with different sequence encoding schemes
 3. Gibbs sampler approach to the prediction of MHC class II binding motifs including pseudo counts and sequences weighting clustering (Hobohm) techniques
 4. Implementation of HMM Baum-Welsh algorithm
 5. Comparative study of PSSM, ANN, SMM for peptide MHC binding
 6. Comparison of “fake” versus “true” cross-validation
 7. Using Theano to optimize ANN learning - deep learning, training methods, regularization, momentum, etc
 8. ...
-

-
- Peptide MHC binding predictions using position specific scoring matrices including pseudo counts and sequence weighting techniques
 - Compare methods for sequence weighting
 - Clustering vs heuristics
 - Benchmark (Peters et al 2006) covering some 20 MHC molecules, compare to best other methods
 - Raw data (SMM) are available at
 - http://tools.iedb.org/main/data/predictions_smm.zip
-

-
- Peptide MHC binding predictions using artificial neural networks with different sequence encoding schemes
 - Benchmark (Peters et al 2006) covering some 20 MHC molecules, compare to best other methods
 - Compare sequence encoding schemes
 - Sparse, Blosum, composition, charge, amino acids size,..
-

Gibbs sampler

- Gibbs sampler approach to the prediction of MHC class II binding motifs
 - Develop Gibbs sampler to prediction of MHC class II binding motifs including cluster-based sequence weighting
 - Benchmark Nielsen et al 2007 covering 14 HLA-DR alleles
 - Compare heuristic versus cluster-based sequence weighting
-

Comparative study

- Compare methods for MHC peptide binding
 - PSSM
 - ANN
 - SMM
 - Data: Benchmark by Peters et al 2006 covering some 20 MHC molecules
-

Method evaluation using cross-validation

- Compare performance of data-driven prediction methods when evaluated using cross-validation
 - What is the difference between the “fake” and “true” cross-validated performance as a function of
 - Model complexity (ANN versus SMM)
 - Data set size
 - ..
 - Data: Benchmark by Peters et al 2006 covering some 20 MHC molecules
-

- Implement Baum-Welsh HMM training
 - Based on code from Tapas Kanungo HMM toolkit
 - Hidden Markov Model (HMM) Software: Implementation of Forward-Backward, Viterbi, and Baum-Welch algorithms.
 - Some links
 - <http://www.kanungo.com/software/umdhmm-v1.02.tar>
 - <http://www.kanungo.com/software/umdhmm-v1.02.zip>
 - <http://www.kanungo.com/software/umdhmm-v1.02.README>
 - <http://www.kanungo.com/software/hmmtut.pdf>
 - Maybe include an MC modelling fitting for comparison
- Test code on un-fair casino example

Tools for ANN training

- Use ANN library tools (Tensor-flow, PyTorch, Keras, ..) to optimize ANN learning/training
 - Number of hidden layers (deep learning),
 - Training methods,
 - regularization,
 - momentum,
 - etc.
 - Data
 - Benchmark (Peters et al 2006) covering some 20 MHC molecules
-