

Sommersemester 2017

## **Computational Immunomics** (BIOINF 3360)

### **Projekt**

Abgabetermin: 30.06.2017, 24:00 Uhr

## **Projektüberblick**

Im Rahmen des Projekts soll eine Methode für MHC-I Bindevorhersage für ein spezifisches MHC Allel entwickelt werden. Nach Projekt-Abgabe werden die Methoden auf einem unabhängigen (streng geheimen) Evaluierungsdatensatz miteinander verglichen und ausgewertet. Alle Unterlagen findet man unter im ILIAS unter *Projekt*.

## **Regeln**

- Das Projekt wird in Teams von 2 bis 3 Personen bearbeitet.
- Trainingsdaten werden auf ILIAS bereitgestellt (*project\_training.txt*)
- Keine Verwendung existierender Implementierungen zur Epitopvorhersage.
- Prediktorperformanz wird mittels ROC-AUC auf einem internen Validierungsdatensatz ermittelt und verglichen

## **Anforderungen**

- Ausführliche und systematische Evaluierung der Methode (Cross-Validation, etc.) mit Hyperparameter Optimierung (Grid-Search)
- Implementierung der Methode (s.u.)
- Schriftlicher Projektbericht (s.u.)
- Mündliche Projektpräsentation (s.u.)

## **Implementierung**

Idealerweise wird eine neue Methode entworfen und implementiert. Falls ihr euch entscheidet eine bereits existierende Methode zu reimplementieren ist eine detaillierte Motivation (zB existierende Implementierung zu langsam, neue Möglichkeiten der Implementierung, etc.) notwendig. Auch sollte eure Implementierung gründlich mit der bereits existierenden Implementierung verglichen werden (Performanz, Laufzeit, etc.) und auch gründlich auf Nachteile und mögliche Erweiterungen eingegangen werden.

Allgemein gilt für die Implementierung:

- Programmierumgebung eurer Wahl (R, Python, C++, Java, Kombinationen etc.). Es bieten sich die entsprechenden Machine Learning Bibliotheken an, z.B. Shogun (C++ mit Bindungen zur weiteren Sprachen, u.a. R und Python), scikit-learn (Python), tensorflow (C++, Python), MLPack (C++), dlib (C++), etc. Allgemein kann man einen entsprechenden Algorithmus auch selbst versuchen zu implementieren, was aber in Bezug auf den zeitlichen Rahmen des Projekts nicht empfohlen ist.
- Ausführliche und sinnvolle Dokumentation
- Das resultierende Programm sollte funktionieren ;)
- Ausführliche Installationsanleitung und möglichst einfacher Installationsmechanismus (also Makefile, setup.py oder Ähnliches).
- Das Programm akzeptiert folgenden Input:

```
< sequence1 >
< sequence2 >
< sequence3 >
...
```

- ... und folgenden Output erzeugen:

```
< sequence1 >< tab >< classification >
< sequence2 >< tab >< classification >
< sequence3 >< tab >< classification >
...
```

- Eine Beispiel-Implementierung findet sich im ILIAS (*example\_predictor.py*). Dort findet man auch Muster für Input- und Output Dateien (*test\_input.txt* und *test\_output.txt*).

## Bericht

Der Bericht sollte 6-8 Seiten umfassen und in einem wissenschaftlichen Stil verfasst sein. Der Bericht sollte folgendermaßen aufgebaut sein:

- Einleitung (Biologische Motivation und Grundlagen)
- Beschreibung der verwendeten Materialien und Methoden
- Ergebnisse der Untersuchung
- Diskussion der Ergebnisse

Im ILIAS findet sich ein tex-template (*Paper\_template.zip*). Overleaf (<https://www.overleaf.com/>) bietet sich zur kollaborativen Erstellung von Tex-Dokumenten an. Dort findet man auch weitere Templates.

## Präsentation

Die Präsentation erfolgt am 04. Juli 2017. Im Anschluss an die 15-minütige Präsentation wird es eine 5-minütige Diskussion geben. Der Vortragsanteil sollte (im Großen und Ganzen) gleichmäßig auf die Team-Mitglieder verteilt sein. Folien sind von Vorteil ;)

## Benotung

In die Benotung fließen ein:

- schriftliche Ausarbeitung (finaler Report, nicht der Zwischenbericht) (40%)
- Implementierung der Methode (40%)
- mündliche Präsentation (20%)

## Zeitplan

- 23. Mai 2017: Projektausgabe
- Allgemeines Mentoring während des Übungstermins. Bei größeren Problemen, Termin vereinbaren.
- 06. Juni 2017, 24:00 Uhr: Kurzer Zwischenbericht (höchstens 1 Seite). Abgabe elektronisch über ILIAS. Geht nicht in die Bewertung mit ein.
- 30. Juni 2017, 24:00 Uhr: Finale Abgabe des Projektes (Bericht, Daten, Code) via ILIAS
- 04. Juli 2017, während des Vorlesungstermins: Projektpräsentationen