

Mendelian disease RNA-seq abnormal gene discovery

Halima Bensmail

Qatar Computing Research Institute

Project Description:

A deep autoencoder is composed of two symmetrical deep-belief networks having four to five shallow layers. One of the networks represents the encoding half of the net and the second network makes up the decoding half. They have more layers than a simple autoencoder and thus can learn more complex features. The layers are restricted Boltzmann machines, the building blocks of deep-belief networks.

The aim of this project is to use a deep autoencoder to learn a low-dimensional unsupervised representation of the data. In the latent space, a separate network is trained to predict which of several mixture components a datapoint will fall into; the mixture model parameters are then updated based on these estimates. The loss function combines the reconstruction loss and an average likelihood for the mixture model, and the whole thing is trained end-to-end.

Specifically, instead of using EM algorithm, which is iterative, time consuming, we look for the best combination of compression estimation networks that maximize the likelihood to minimize the sample energy and this requires knowledge in statistical inferences.

Available data are RNA-seq counts available (Kremer data 1 and Murdock data 2) which are available and curated.

Project Type: Research

Duties/Activities:

- a- Implement “deep autoencoder” machine learning to detect unusual genes
- b- Publish a bioinformatics tool at least at the Bioinformatics Journal.

Required Skills:

Knowledge in any kind of programming such as Python or R or others.

Internship Batch: Batch 1 from May 7 to June 29

Mentors

Halima Bensmail

email: hbensmail@hbku.edu.qa