Rajan Kapoor

rkapr.analytics@gmail.com

rkapr.github.io
linkedin.com/in/rkapr

979-398-2987
github.com/rkapr

EDUCATION:

Texas A&M University, College Station, TX
PhD in Electrical and Computer Engineering (GPA: 3.82)   Dec 2021

Texas A&M University, College Station, TX
Master of Science in Electrical and Computer Engineering (GPA: 3.75)   May 2017

Indian Institute of Technology, Patna, India
Bachelor of Technology in Electrical Engineering (GPA: 8.57/10)   May 2014

PROFESSIONAL EXPERIENCE:

Software Intern, Rockwell Collins, Hyderabad, India   May 2013 - Aug 2013

- Implemented MATLAB based OFDM transceiver on Rockwell software defined radio hardware module

RESEARCH PROJECTS/EXPERIENCE:

Understanding nitrogen transport pathways related to grain protein content in wheat

- Identifying gene regulatory pathways that transport nitrogen from senescing leaves to grain with goal of understanding mechanisms that affecting grain protein content in wheat
- Performed WGCNA based gene clustering using R scripts on computing cluster
- Developed shell based optimized script to download large alignment files using globus command line
- Wrote shell script to extract alignments corresponding to predicted UMAMI genes using GTF file
- Performed exploratory analysis of UMAMI gene expression in Arabidopsis to verify orthologs and identify phylogenetically conserved expression patterns
- *Key achievement:* The analysis discovered two modules involved in activating storage molecules including starch, lipids and protein, discovered multiple transcription factors as possible key candidate genes

Gaussian Graphical model with fused lasso penalty for learning causal transcriptional regulations of Sorgoleone biosynthesis genes in Sorghum

- Inferring transcriptional regulation of sorgoleone biosynthesis to understand herbivory resistance in sorghum roots
- Performed Gaussian graphical modeling with fused penalty to uncover causal gene interactions
- Verified the algorithm was able to extract causal gene interactions for sorghum circadian genes
- Wrote scripts for cleaning, mapping, counting, differential analysis of RNAseq data
- Used root and leaf development data from stay green sorghum to identify regulators
- *Key achievement:* Identified 21 potential regulators of sorgoleone biosynthesis genes, verified TF-gene interactions using motifs from plant transcription factor database and comparative genomics

Gaussian Mixture Modeling for Cancer Heterogeneity

- Estimated proportion of different subpopulations in heterogeneous cancer data using expectation maximization with k-means initialization
- Utilized drug target interaction and gene regulatory information from literature for different subpopulations by modelling them as a gaussian mixture of different Boolean networks
- *Key achievement:* The model accuracy was validated on two different in-house datasets of heterogeneous cancer populations.

COURSE PROJECTS:

Mixture of Poisson's for modeling number of daily deaths   Spring 2018

- Mixture of Poisson's to model total number of deaths as mixture of summer and winter deaths
- Implemented and compared convergence of gradient descent and Newton Raphson optimization for maximum likelihood estimates
- Calculated *sympy* expressions for gradient and hessian in Python, then converted to *numpy* functions to avoid errors due to hard coding expressions

*Zeroinfl***: Zero Inflated Poisson regression (ZIP) in Python** (with Eric Chuu)   Spring 2018

- Modeled a process switching between perfect state with no errors and imperfect one with Poisson distributed errors using ZIP regression based on *pscl* (R) and *statsmodels* (python) packages
- Wrote functions for likelihood, log-likelihood, gradient and maximum likelihood estimation using BFGS optimization, with expectation maximization-based initialization for different link functions
- Extended functionality by including zero inflated negative binomial and zero inflated geometric regressions
- *Key achievement:* Provided fully functional zero inflated regression functionality in Python

Adaboost based face detection using Voila Jones framework   Fall 2019

- Implemented five stage Adaboost classifier in Python with decision stump (one step decision tree) as weak learner for detecting face images using vertical, horizontal and diagonal Haar features
- Vectorized code to speed-up repeated processing, observed effects of FP/FN penalty on empirical error
- *Key achievement:* Achieved **3.5x improvement** in training time at cost of 2.25% reduction in empirical accuracy by filtering robust features

Minimum description length (MDL) based Boolean network learning   Spring 2020

- Implemented MDL based gene network learning algorithm from short Boolean time series by minimizing sum of error and model-description entropies
- Tested and verified the code on Boolean time series data generated using *BoolNet* package in R
- *Key achievement:* Released MATLAB code based on Dougherty et al. paper was for open source use

SKILLS:

- R, Python, shell scripting (SLURM, LSF), MATLAB, SQL, Jupyter/R Notebooks, C/C++

RELEVANT COURSES/ CERTIFICATIONS:

- *Engineering/Statistics:* Regression Analysis (A), Statistical computing in R & Python (A), Applied Statistics & Data Analysis (B), Pattern Recognition (A), Distribution Theory (A), Information Theory (A)
- *Biology/Bioinformatics:* Bioinformatics (S), Bioinformatics Command Line (A)
- *Online Certifications:* Introduction to SQL, Learn the Command Line, Algorithms & Data Structures

PUBLICATIONS:

- A Gaussian Mixture-Model Exploiting Pathway Knowledge for Dissecting Cancer Heterogeneity, IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2018

HONORS/ AWARDS:

- Texas Engineering Experiment Station (TEES) Research Assistantship, Sept 2015 - present
- Texas International Student Scholarship, Fall 2015, Spring 2016