Data Scientist for Machine Learning Applications in Medical Care

Job description: The AI in Medicine program at UCLA, led by Professor Eran Halperin, aims to develop algorithms for the prediction of clinical outcomes from medical records, physiological waveforms, imaging data, and genomic data. In order to achieve this goal, novel and creative machine learning and statistical analysis is needed. The candidate will be in charge of developing and implementing such algorithms on a variety of data sources and on a variety of clinical outcomes.

Work environment: Our projects involve the collaboration of interdisciplinary teams including machine learning researchers, computational biologists, and clinicians. The candidate will be expected to be a good communicator, particularly with the ability to communicate in an interdisciplinary environment, where different individuals have different expertise. The work is highly practical and translational - the output of the work is improved diagnostic computational tools using existing data that can be incorporated in UCLA Health and beyond.

Job requirements:
- The candidate needs to have undergraduate degree in either computer science, statistics, engineering, biomedical informatics, or a related field.
- A PhD degree in a related field is an advantage (but is not necessary).
- Programming competence demonstrated in at least one or more of these programming languages: Python, R, Java, C++, Matlab.
- Knowledge and deployment of advanced statistical and machine learning concepts used in big data analysis including nonparametric tests, ANOVA, mixed models, modern supervised and unsupervised machine learning algorithms such as SVM, Random Forest, PCA, Clustering, and Neural Networks.
- Ability to communicate clearly in an interdisciplinary environment.
- Knowledge and experience working with genomic data or with medical records data is an advantage but it is not necessary.
- Ability to work in a Linux environment - advantage
- Software tool development experience: source control (git), packaging, documentation - advantage.

Requirement self-testing: In order to have an idea whether you are qualified for this work, please see if you can answer the following questions:

1. Explain what is a Bonferroni correction for multiple hypotheses.
2. Let X,Y be two independent standard Normal random variables. What is the expectation of X*Y? What is the variance of X*Y? Is X*Y also normally distributed?
3. You have an array of 1,000,000,000 numbers. Find an efficient algorithm that returns the 100th largest number of the array.
4. If v is an eigenvector of X as well as an eigenvector of Y, is it also an eigenvector of X+Y?
5. In a random binary sequence of length n, where each bit is independently picked to have value 0 or 1 with probability 0.5, what is the expected number of substrings "000"?
6. In a machine learning problem briefly explain what are the following: test error, training error, overfitting, cross-validation.

If the above questions are easy for you, and if you are interested in improving medical care using machine learning, please send your resume and a brief cover letter email to ehalperin@cs.ucla.edu.