Assignment – Continuous Optimization and GLM’s

# Questions:

## Continuous Optimization

1. What does the first order derivative tell us about a function?
2. What does the second order derivative tell us about a function?
3. What is the Jacobian matrix?
4. What is the Hessian matrix?
5. Does iteratively reweighted least squares use only first, or first and second order derivatives?
6. Why does IRLS tend to converge faster than gradient descent for a GLM?
7. What is the difference between gradient descent and stochastic gradient descent?
8. Is stochastic gradient descent more efficient than gradient descent in general? Why or why not?
9. What is the difference between coordinate descent and gradient descent?
10. Why does IRLS scale poorly with increasing number of predictors?
11. If I have a large number of predictors (1000’s) and want to distribute my calculations using parallel frameworks like H2O or Dask or Apache Spark, should I use IRLS or L-BFGS? Why?

## GLMS

1. How does a general linear model differ from a generalized linear model?
2. What are the three components of a generalized linear model?
3. What optimization method is typically used for finding solutions to generalized linear models?
4. What is maximum likelihood and what does it have to do with GLMs?
5. What is the canonical (natural) link for the binomial distribution?
6. What is the canonical (natural) link for the Poisson distribution?
7. What is the canonical (natural) link for the Normal distribution?
8. Why is a binomial family with a logit link (logistic regression) sometimes easier to fit than a binomial family with a log link (log-binomial regression)?