**REI602M** **Quiz 22.1.2021**

Hand in on Gradescope before 22:00 on Jan. 23 (Saturday). Each question will be given 1, 0.5 or 0 points as follows. If the question is more or less correct it gets 1 point. If it is partly correct it gets 0.5, and if it is missing or completely wrong it gets 0 points.

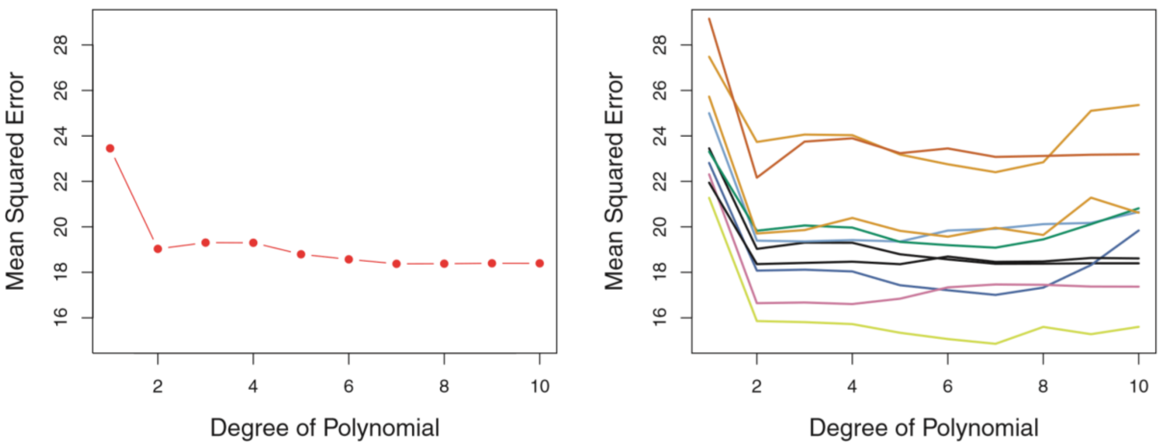
1) Explain in 1 – 2 sentences why it is important to scale input variables prior to training a model (estimating its parameters from data).

**Svar:**

If an input variable is very large and/or has a large spread of values it may result the error gradient values to be very large, therefore the weight values will change dramatically.

2) You are to construct a predictive model $f_\theta(\boldsymbol{x})$ of drug activity as a function of dosage, formulation and other variables. You have a data set with 88 measurements. The model is a polynomial regression model (linear in $\theta$, nonlinear in $x$). The data set is split randomly into a training set (75%) and test set (25%). The left figure below shows the mean squared error (MSE) for polynomials of degree 1 to 10 for one such split. Since the test set is small (22 points) we can’t be sure that it provides an accurate estimate of the true error. Therefore several random train/test splits are performed and the MSE calculated for each test set (right figure below).

Which polynomial degree do you suggest and why?




Svar:

I would choose polynominal degree between 2 and 4 since the elbow of the graph is at 2 for every combination of train/test sets.

3) [From past exams] For each of the statements below, mark it as TRUE or FALSE. Provide a 1 – 2 sentence explanation of your answer.

a) It is computationally more expensive to compute the prediction of a point **x** with a logistic regression classifier, compared to a k-nearest neighbor classifier.

True or False? Why?

**Svar:**

False because KNN is comparatively slower than Logistic Regression

b) Þar sem flokkun er sértilvik af aðhvarfsgreiningu (e. regression) þá er "logistic regression" sértilfelli af línulegri aðhvarfsgreiningu.

**Svar:**

True, for the case of y as categorical.

b) Since classification (the y’s are discrete) is a special case of regression (the y’s are continuous) therefore logistic regression is a special case of linear regression.

True or False? Why?

**Svar:**

False in the case of y as discrete, it would be a special case of linear regression if the target variable was categorical in nature.

4) The sigmoid function is defined as g(z) = 1 / (1 + exp(-z)). Show that the derivative of the function can be written as g'(z) = g(z)(1-g(z)) [which shows that computing the derivative is almost free, once g(z) has been computed] *Hint*: (1 / f(z))' = (-f' / f(z)$^2$).

$$g(z) = \frac{1}{1 + \mathrm{e}^{-z}}$$

$$let's\ use\ the\ Quotient\ rule: f(x) = \frac{g(x)}{h(x)} and\ therefore$$

$$f'(x) = \frac{g'(x)h(x) - g(x)h'(x)}{|h(x)|^2}$$

therefore $g'(z) = \frac{0*(1+e^{-z}) - 1*(-e^{-z})}{(1+e^{-z})^2} = \frac{e^{-z}}{(1+e^{-z})^2} = \frac{1}{e^{z}(1+e^{-z})^2}$

if we expand the resault we get:

$$\frac{1}{e^{z}(1+e^{-z})^2} = \frac{1}{e^{z}(1+e^{-z})^2} = \frac{1}{e^{z}(1+e^{-2z})} = \frac{1}{e^{z} + e^{-z}} = \frac{1}{1 + e^{z}} * \left(1 - \frac{1}{1 + e^{z}}\right)$$

therefore $g'(z) = g(z) * (1-g(z))$