Vincent Nguyen

ORIE 4740 Homework 4

TA: Jian Wu

Sect: Monday 2:30

1)

Multiple linear regression (that does not include interaction terms) cannot capture the situation where a woman is more likely to contribute if she has pets than if she doesn't, while a man is less likely to contribute than if he doesn't. This is because such a statement implies that there are two different linear relationships between the pet ownership (the predictor) and monetary contribution (prediction). Namely, one situation (the situation that the veteran is a woman) assumes that there is a positive linearity between the ownership of a pet and likelihood of donation, where as another situation (the situation that the veteran is a male) assumes that there is negative linearity between the ownership of the pet and likelihood of donation. Multiple linear regression can only capture one case of linearity between the predictor and the outcome (either positive linearity or negative linearity) and so it cannot capture the situation in question.

This is because, when looking at the traditional model of multiple linear regression in which Y = f(X), f() is a linear function such that Y = B0 + B1\*x1 + B2\*x2... Since the traditional multiple linear regression model assumes that B1 and B2 are constant scalar values, they can only express one kind of linearity (as opposed to the two kinds of linearity that the situation in question requires – positive and negative linearity)

However, the situation can be captured if we included interactions terms, such as Y i.e Y = B0 + B1\*x1 + B2\*x2 + B3x1\*x2.

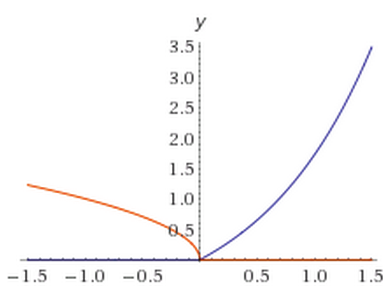
2)

Let u(t) be step-function such that u(t) = 1 for t>0 and u(t) = 0 for t<=0. An error measure that captures overestimation as being much more costly than underestimation would be as follows:



where hat{\theta} is an estimator (that is some function of the predictors) with respect to an estimated parameter theta. An illustrative plot for the error measure for such a function is shown below as a function of the difference between the estimator and the estimated parameter i.e





As you can see, the error measure grows exponentially as the estimator overestimates the estimated parameter. However, as the estimator underestimates the error measure grows like (-x)^1/2. The only problem with this predictor however is that at smaller values, the error measure for underestimation will exceed that of overestimation (when the absolute value of the error is less than 0.5).

An example of an error measure that does not suffer from this is as follows



such that a<b. This error measure will grow linearly for both over and underestimation, but will grow at a higher rate for over estimation. Furthermore, overestimation by a certain amount will always have a higher error than for underestimation by said amount. This means that the cost associated with overestimation with this error measure is always more than the underestimation by the same number of units.

An example where overestimation is more costly than underestimation is in evaluating the likelihood of winning if playing a black jack round. Assuming that every dollar that you continue to earn at the table is less and less valuable (because the negative utility of being bankrupt is not worth the prospect of doubling, tripling or quadrupling your current wealth), this means that your threshold at which you decide to play becomes steadily lower as you become richer. In other words, as you become richer, it becomes a worse idea to gamble your money because the kth dollar is always more valuable than the k+1th dollar. This is the same as saying 1 dollar to you if you are poor is a lot more valuable the 1 dollar if you are rich.

In light of this, overestimating whether or not you should play a blackjack round is more costly than underestimating if you should play a blackjack hand because the money that you lose is a lot more costly than the money that you should gain: in other words the expected payoff of overestimating is generally worse than the expected payoff of underestimating your winning probability and thus not playing – which has a payoff of 0).

3.

With the restriction in the question, there are still 2^4 different models (16 different models). This is because there are 4 different predictors that we can choose to include in each model, which gives us a total of 2^4 different models (different in the sense that they include different sets of predictors).

4

a)

There are 5 parameters:

Y = B0 + B1x1 + B2x2+ B3x3 + B4x4

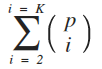
All 5 Betas (B0,B1,B2,B3,B4) must be estimated

b)

To count the interaction terms, we would calculate 4 choose 2 i.e how many unique sets of 2 can we choose from a set of 4. That is 6 terms. Thus, multiple linear regression that also includes the pairwise interaction terms would have 11 terms total (5 from standard regression, 6 from pairwise interaction terms).

c)

if we were to include pairwise, three-way, 4 way, all the way up to K-way interactions, we would have to include



extra terms (not including the original terms from standard multiple linear regression).

The problem with including all of these terms is that computationally impractical to estimate all of the terms in this fashion, because the number of terms that we need to calculate grows very quickly if the number of predictors is large.

6)

Let 'p' be the probability that a customer will not default on his loan and so (1-p) is the probability that the customer will default on his loan. Let b be the benefit to the bank (presumably in positie dollars), while c is the cost to the bank (c is presumably a negative value in dollars). As long as the following condition holds

pb + (1-p)c >0

* p > c/(c-b)

then the bank should accept the grant the loan (because the expected payoff is greater than 0 (which is greater than the payoff than turning down a loan application)). Which is to say, that the threshold to surpass is c/(c-b).

7)

Value of Z1 : v = -0.2+0.12(5)+0.35(10) = 3.9

σ(v) = 0.98

Value of Z2 : v = -0.1+0.6(5)-0.5(10) = -2.1

σ(v) = 0.109

Value of Y

1.2 + 3(0.98) + 4(0.109) = 4.576

σ(Y) = 4.576

therefore the value of Y = 4.576