**Introduction to Data Mining**

***Quiz 1, Spring 2016 Total Marks = 15;*** *Time: 15 min.*

*Declaration:* By submitting this quiz for grading, I affirm that I have neither given nor received help from another examinee and acknowledge that **this is a closed-book, closed-notes test.**

Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Signature: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Q1. (2+2+2 pts)** Explain briefly (e.g. by giving an example) what you understand by

1. an *ordinal* variable

An ordinal variable is a categorical variable with clear ordering of the categories. For example, you can have an ordinal variable for monthly income with three categories (low, medium, high).

1. obtaining a *marginal* pdf from a multivariate probability density function (pdf).

If two random variables X and Y have a joint probability density function f(x,y), then the marginal pdf of X, denoted by fX(x), can be obtained by ∑yf(x,y) (for discrete case), or by (for continuous case). This is basically the probability that X will take on a value x (Pr(X=x)). Can generalize to joint distribution over multiple r.v.s

1. a boxplot

(from homework! Or see Wikipedia)

curse of dimensionality: see text

**Q2. (2+3 pts).** How does using Huber loss instead of squared loss make linear regression more resistant to outliers? Mathematically show how using the Huber loss function can be viewed as solving for a weighted version of standard linear regression that is based on minimizing the “sum of squared errors” loss function. (with the appropriate choice of weights, the solutions will be identical).

Standard linear regression uses squared loss (r2, where residual r = y – y’) for all data points, whereas huber loss uses squared error function for smaller residuals, and linear error function (|r|) for larger residuals. So we can view linear regression with huber loss as a weighted version of standard linear regression, where weights for the smaller residuals are 1, and weights for the larger residuals are proportional to 1/|r| so that the error function becomes linear. (see lecture slides for more details)

**Q3 (4pts).** Briefly, what are the assumptions regarding

1. the relation between the dependent variable (Y) and the independent variables (Xs)
2. the nature of the added noise

made in the standard (multiple) linear regression model?

a. conditional mean is E[Y | 𝛃, X] = XT 𝛃, conditional mean of Y (given 𝛃 and **X**) is linear in X

b. The residuals 𝜺i  are independent and identically distributed (constant variance, 0 mean normally distributed) for all i.