**Probability basics**

1. Bayes theorem:
2. How many parameters does a Gaussian distribution have ? How do they affect the probability density function of the distribution ?
3. Explain Strong law of large numbers in your own words.
4. Explain Central limit theorem in your own words.

**Linear Regression**

1. Explain the optimization problem you solve to infer the weights/coefficients in a linear regression model.
2. How many minimum number of data points do you need to fit a 7*th* degree polynomial.
3. What is the minimum number of data points you need to fit a straight line model
4. If you have more than 4 data points, you can always fit a cubic polynomial to the data. True or False ? Give your reasoning.
5. If we have 4 data points (*x*1*, y*1), (*x*2*, y*2), (*x*3*, y*3) and (*x*4*, y*4) and we want to fit a cubic polynomial, write down the dimensions of the matrices X, Y and *β*, where

*β* = (*XT X*)*−*1*XT Y*

1. In part 5 above, using the cubic polynomial model fitted on 4 data points, can we predict y values at more than 4 new x locations ? Why or why not?
2. Explain a situation in which *β* computation in (1) will produce an error.
3. For the following models mention if they are linear or not and give your reasoning. In case they are linear, mention the dimension of X matrix (assuming 1000 data points) you would formulate to compute *β* vector
   * *y* = *β*0*x* + *β*1
   * *y* = *β*0*x* + *sin*(*x*)
   * *y* = *β*0 + *β*1 + *β*2*cos*(2*x*)
   * *y* = *β*0 + *β*1*sin*(*x*)
   * *y* = *β*0 + *β*1*sin*(*x*2) + *β*2*cos*(*x*)
   * *y* = *β*0 + *β*1*sin*(*β*2*x*)
   * *y* = *β*0 + *β*1*sin*(*log*(*x*))
   * *y* = *β*0 + *β*1*ex*
   * *y* = *β*0 + *eβ*1 *x*
   * *y* = *ββ*1 + *β*2*x*

0

9.Mention the 4 conditions that make a regression model linear. When is a regression model non-linear?

**Model Selection**

1. Explain what is hypothesis testing.
2. t-test for quantifying the significance of individual coefficients in linear regression model is same as examining confidence intervals for these coefficients. True or False ? Explain.
3. If *p − value <* 0*.*05, what does that tell you about the Null hypothesis.
4. Explain the terms in the F-statistic formula. 5. Should we accept or reject the Null when F is large?

**Regression diagnostics**

1. If you have 4 data points (*x*1*, y*1), (*x*2*, y*2), (*x*3*, y*3) and (*x*4*, y*4) and we want to fit a cubic polynomial, what would be the dimension of the hat matrix *H* = *X*(*XT X*)*−*1*XT* . What would be the dimension of *H*5 ?
2. H is a linear operator that transforms observations Y to predictions *Y*ˆ . True or False ?
3. How can we express the residual vector in terms of the hat matrix ?
4. How do you compute the Cooke’s distance ? Explain each term in the formula.
5. How do you compute the *kth* lag autocorrelation ? Explain each term in the formula.

**Confidence Intervals**

1. Explain the statement: Estimated parameter set *β*ˆ is an unbiased estimate of the true param- eter set *β* in the ordinary least square formulation.
2. How can you use the confidence intervals computed for individual *βi* ?
3. If you fit a 10*th* degree polynomial, which coefficient out of *β*0 to *β*10 you expect to be most concentrated around 0 and why ?
4. Explain confidence intervals in reference to the mean response.
5. Explain prediction intervals in reference to the new observations.
6. Explain why prediction interval is thicker than confidence interval

**Ridge Regression**

1. Enumerate the differences between Ridge Regression and Linear Regression.
2. Explain the acronym BLUE with respect to ridge regression.
3. What does the constraint *||β||*2 *≤ c*2 achieve in Ridge Regression formulation.

2

1. In the estimate:

*β*ˆ*ridge* = (*XT X* + *nλI*)*−*1*XT Y*

does the model overfit or underfit when you increase *λ* beyond the optimal value. How about when you make *λ* very small.

1. The distance of Ridge Regression estimate from origin is always less than distance of OLS estimate from origin. True or False ? Give your reasoning.
2. The computational expense of Ridge Regression is more than Linear Regression. True or False? Reasoning?

**Lasso Regression**

1. Explain the significance of *|| · ||*1 penalty in the Lasso Regression formulation.
2. How does Lasso Regression algorithm lead to feature selection.
3. Lasso Regression produces a simpler model than Ridge Regression. True or False ? Explain.
4. When would you prefer Lasso Regression over Linear Regression and Ridge Regression.
5. Compute *||β||*1 and *||β||*2
6. For very noisy data, would you prefer Linear Regression/Lasso Regreesion/Ridge Regression and why