**QUESTIONS**

1. **What is the autocorrelation function and how it may be used?**

The autocorrelation function allows us to detect the linear relationship between lagged values of a time series. The idea is that it is possible that consecutive observations in a time series could be autocorrelated, so in order to spot this behavior we use the correlogram. The correlogram is a tool which provide a plot of the autocorrelation between observation, for example the first bar describes the autocorrelation between observation distant one step). This tool provides also an interval that determine the significance or not of the bar. Analyze the correlogram of observations in the data is important in order to find a trend or a seasonality but it also useful for analyzing the residuals in order to understand if our model has been able to describe well the data. It is a diagnostic tool.

1. **Describe the multiple linear regression model.**

The multiple linear regression model is composed by:

* a response variable Y and
* p predictors each one of them multiplied by the correspondent coefficient which quantify the relationship between the predictor and the response.
* An error term.

The estimation of the parameters is done by minimizing the least square. We can evaluate the performance of the model using the R square statistic, and then the significance of the parameters through the F-statistic

\*F test is for global significance of model for significance of parameter we use t=bj/se(bj)

1. **Describe the problem of multicollinearity in a linear regression model and explain how it may be detected.**

The multicollinearity problem arises when ~~two~~ (when we say multicollinearity it means three or more) or more predictors are correlated to one another. With this problem we could have a bad effect on the accuracy of the estimation of our parameters or a higher standard error for the predictor and consequently we may fail to reject the null hypothesis. The simpler way to detect the collinearity is to look at correlogram matrix of the predictors but the best way is to compute the VIF which if it is greater than 4 indicates that there is multicollinearity problem.

1. **Describe the multiple linear regression model applied to time series.**

The multiple linear regression model can be used with monthly, quarterly or yearly data. The trend is modeled by a predictor and the seasonality through s-1 dummy variables. The seasonal parts produce straight lines with different intercept and same slope.

1. **Describe autocorrelated residuals and explain how they may be detected.**

The autocorrelated residuals indicates that our model does not captured and modeled the data in a good way and probably we have to choose a different model or add variables. This kind of problem can be detected by analyzing the correlogram, if there are significant bar we have positive/negative autocorrelations between lagged residuals. Another way is to compute the Durbin-Watson test, but is useful only when there is an autocorrelation between lag 1 residuals.

1. **Describe the Bass Model.**

The Bass Model is a non linear model used to describe diffusion process of a product, technology ecc. The model gives the instantaneous sales at time t and it is characterized by the parameters:

* P which represents the innovation
* Q which represents the imitation
* M represents the maximum market potential and is constant.

This model is good to infer market qualities that a priori we don’t know (the p and q parameters) but it has some limitations: the assumption that m is constant, the BM does not take into account marketing strategies and it is a model for products with a limited life cycle.

1. **Describe the Generalize Bass Model**

The generalize bass model is a generalization of the BM, it adds an intervention function x(t) that describe the type of shock that occurs during the diffusion process. If x(t) = 1 we have the BM, if 0<x(t)<1 the shock blocks the diffusion although if x(t) > 1 the shock accelerates the diffusion. We can have different type of shock. The positive exponential shock is suitable for identifying the effect of strategies with the aim of increase the diffusion but also for analyze the consequence of a competitor in the case of negative exponential shock. The rectangular shock is good to identify the effect of strategies which are applied during an interval of time. The important characteristic of the GBM is that it can be used as diagnosing tool, for instance we want to understand from the time series, the effect of a marketing campaign. In other words we want to describe in a retrospective way if what is happened is a real shock or just noise.

1. **Describe the GGM**

The GGM model is a generalization of the Bass Model and it consider the market potential as function of time (not constant as in BM) and a communication process. It consists in a scale parameter that multiply the communication process and the adoption phase. In addition to the parameters of the GBM there are 2 more parameters that describe the process of communication, namely pc and qc.

1. **What are the elements to consider in evaluating the performance of an innovation diffusion model?**

We can consider different elements to evaluate the performance of an innovation diffusion model: the first one is the goodness of fit respect our data (we can see it with a cumulative plot or a simple plot), then we can see the estimations of our parameters: m (market potential), p (innovation), q (imitation) (with GBM we add also the parameters in the intervention function x(t), and in the case of GGM all the parameters that describe the communication part (K, p\_c and q\_c)) can be evaluated trough their significance. Another element to evaluate them, is the residuals. In particular, seeing the plot (to detect visible trend or seasonality not captured) or correlograms (ACF and PACF) to see possible autocorrelations. Finally, we can make some qualitative evaluations seeing the behavior in the instantaneous plot: does our model capture also particular changing in a particular region (changing in policies, market campaign ecc.)?

1. **Describe ARIMA models.**

ARIMA models are useful approach for time series forecasting. The model combines differencing with autoregression and moving average model. The differencing part allow us to remove the seasonality/trend of a time series in order to obtain a stationary time series. Usually the ARIMA model is defined by three parameters (p,d,q): p refers to Autoregressive part and indicates the number of lagged values of y used, d is the degree of differencing and q refers to the Moving Average part and is the number of errors used.

An extension of ARIMA model considers the seasonality and it is composed by two part: the first is the non-seasonal part (p,d,q) and the second is the seasonal part (P,D,Q)s.

1. **Describe the bias-variance trade-off.**

The bias-variance trade-off is a problem that is associated with model selection. Ideally, we want to choose a model that is able to capture the patterns in the data but also it can generalize well unseen patterns.

High-variance learning methods may be able to represent their training set well but are at risk of overfitting to noisy or unrepresentative training data. In contrast, algorithms with high bias typically produce simpler models that may fail to capture important regularities (i.e. underfit) in the data.

1. **Describe the KNN for a regression problem**

The KNN is a non parametric model which can be used in regression problem. The idea is based on the fact that we want estimate the regression function f(x) which corresponds to the expectation of y given x0, this represent the minimal solution for the quadratic loss function. The formulation of the KNN defines a set of k observations closed to x0 after computing an appropriate distant measure. In the end the function estimation for x0 is the expectation of y given the observation in the neighbourd set.

1. **Describe advantages and drawbacks of the KNN approach for a regression problem.**

The KNN approach is a non-parametric approach that in some cases can perform in better way respect a parametric model. For example, it is a valid tool respect the simple linear model when the relation between x and y shows a non-linear behavior. The exploitation of the neighbors allows a good flexibility to the model. The main drawbacks arises when we deal with more predictors (curse of dimensionality). In many dimensions the distances between observation doesn't allow a good description of the data. Finally, we have to choose the right number of neighbors considering the bias-variance trade-off: with few neighbors we have high variance and low bias, while with many neighbors high bias and low variance (extreme case: k = n, that means horizontal line with y equals to the mean of y).

1. **Describe Local Regression**

Local regression is based on the idea that we have an interpolation of the data, we estimate a simple linear regression at each point x0 based on a certain number of neighbors points, given the shape of Wh. To each neighbor xi we assign a weight: if xi and x0 are close the weight is high otherwise is low. We have two parameters: the W is a density function called kernel and h indicates the smoothness, if h is big the model consider many neighbors.

1. **Describe loess**

Loess regression is based on the same idea of local regression, so we estimate a simple linear regression at each point x0 considering a certain number of neighbors point given the shape of Wh. The W is the density function called kernel and h indicates the smoothness, in the case of the loess this parameters is variable. The idea is that we take into account the fact associated to the sparseness of data, we computed the value of h according to the sparsity of data, defining the fraction of observation for estimating f(x) at a certain point x0.

1. **Describe splines**

Spline is used to approximate functions or to interpolate data. The idea is to build a function that passes through K points called knots and it is free in the other points. We consider only the knots from 2nd to k-1 and in the interval between two knots the function is a polynomial of d degree of freedom. The obtained function is called regression spline and is a linear combination of basis function. It is possible to consider a variation called ‘smoothing spline’ which adds a penalization parameter (lambda) that perform a smoothing on the function. More lambda is large, smoother is the function. In this case the number of knots coincides with the number of samples indeed if lambda = 0 we have a perfect interpolation with the data.

1. **What is the role of effective degrees of freedom in a nonparametric regression context?**

The effective degreed of freedom allow us to perform some important evaluations like in the context of parametric context. In particular, we can perform an analysis of variance (ANOVA) and calculating the F-score. We can compare different models, for example nested and total, or different nested models, like in the parametric context.

1. **Describe the GAM**

Generalized additive models are the non-linear version of the multiple linear regression. With these models we can use local regression and splines as building blocks. With GAMs we have the same structure of multiple linear regression but instead having simple parameters that multiply predictors, we have functions of predictors. These function are estimated through the backfitting algorithm. The important properties refers to the facts that the non-linear fit could make accurate predictions for the response variable and we can still analyze the effect of each predictor on the response variable.

1. **Describe gradient boosting**

The gradient boosting is a powerful machine learning algorithms which can be used in classification and regression problems. In the case of regression problem, it is based on the idea of improving the prediction given by a model, fitting a regression tree on the residuals. Hence the new prediction will be: y = f(x) + h(x). We can repeat this procedure M times and improve the prediction at each iteration. The gradient boosting is related to the gradient descent because considering the case of quadratic loss function, the negative gradient of the loss is equal to the residuals so basically the gb algorithm fits a regression tree on the negative gradient.

1. **What are the key elements to estimate a Gradient Boosting model?**

A Gradient Boosting model can be estimate with the following ingredients:

• A training set T = f(x1; y1); : : : ; (xn; yn)g;

• A loss function L(y; ^ f);

• A number of iterations for the gradient descent algorithm, M;

The GD algorithm minimizes the loss function through subsequent decrements via the usage of the gradient information evaluated at each iteration.

1. **What the key elements to evaluate the performance of a Gradient Boosting model?**

The elements to evaluate the performance of a GB model is to use different loss function such as the MAE, MSE (even if it’s affected by outliers) and the hubel one.

1. **What is the relative importance in the GBM model**

The relative importance indicates what are the variables that play a major role in terms of producing the prediction of the response variable. In other words: how much a explanatory variable contributes to reduce the prediction error.

1. **What are partial dependence plots?**

These plots are useful to explore better the relationship between our response and explanatory variables accounting for the presence of all the other variables. This is not an indicator of a pure relationship between x and y.

Instead using the GAM, the plots are interpretated like the parameters in a linear regression model: the net effect of a variable on the response variable.