Q1. What is the business problem in this case and how is this business problem converted into an analytics problem?

**Answer**: MHE employed manual system to take patient’s feedbacks which account for around 30% coverage of total registrations in month. This delayed the real time implementation of patient’s feedback and updating of action taken back to patients.

A method employed to convert manual feedback into real time feedback gives tangibility to process and an opportunity to deploy analytics solutions to data gathered. Now the problem is much able to be viewed from lens of analysis of data gathered and can be solved from predicting the net promoter score.

Q2. What is the extent of missing information is survey data? What implication could it have on model building? Use KNN to ﬁll in missing information.

Not to be attempted as per Instructor.

**Q3**. How can we estimate sensitivity and speciﬁcity of three-class problem? Provide the formulas.

**Answer**: For a multi class classification problem, we can deploy one vs all approach to find sensitivity and specificity. For example, consider following matrix. In this problem we use referencing method to determine sensitivity and specificity.

Prediction Detractor Passive Promotor

Detractor 21 9 1

Passive 11 50 26

Promotor 12 58 176

In this example sensitivity of Detractor is TP(Detractor)/(TP(Detractor)+FN(Detractor) where

TP(Detractor) = 21

FN(Detractor)= E (detractor and passive) +E(Detractor + promoter)

FN(Detractor) = 9+1 =10

So, sensitivity = 21/31 = .67

Similarly, Specificity is TN(Detractor)/TN(Detractor)+FP(Detractor)

TN(Detractor) = 50+176+58+26= 310

FP(Detractor) = E(Passive and Detractor) + E(Promoter and Detractor)

FP(Detractor) = 11+12 = 23

Specificity = 310/(310+23) = .93

**Q4.** What is quasi-complete separation? Which variables in the Manipal Hospital dataset are leading to quasi-complete separation?

**Answer**: Quasi complete separation leads to the inability of a predictor variable to separate response variable in clear cut classes. Or it can separate predictor class up to some extent.

In following example, we can see that for P1(12) we have y as 0 and 1. While for P1 < 12 and P1 > 12 Y is 0 and 1 respectively. It means that for all P1= 12 we have achieved quasi separation in data.

Y P1 P2

0 10 2

0 10 0

0 12 -3

1 12 2

1 14 1

1 15 8

1 17 1

1 18 4

1 19 2

1 20 4

We come to know about this situation in Data set when we get error message like “fitted probabilities numerically 0 or 1 occurred” in R.

We can keep or remove variables from data as desired to build the model.

When we encountered this problem in our Data, we found out following variables causing quasi and complete separation.

Marital Status Country EM\_NURSING Bed Category State

EM\_DOCTOR DOC\_ATTITUDE NS\_NURSESATTITUDE OVS\_OVERALLSTAFFATTITUDE

CE\_NPS.

**Q5**. What is orthogonal polynomial coding and how is it implemented in contrasting ordinal variables?

**Answer**: When using regression, categorical variables of k categories are pass as k-1 sequence of variables. Regression coefficients of these k-1 set corresponds to set of linear hypotheses. To understand this phenomenon, we will have to dive deeper in vector and its projections on planes.

Multiple Regression assumes that there is no collinearity among variables but, it is not the case. Orthogonal polynomial regression helps treating its vectors in an Orthogonal plane to keep relationship between them independent.

90-degree angle between two vectors in plane illustrate independent relation between them. If angle decreases from 90 degree, they are positively related and vice versa. To handle collinearity among variables we try to keep plane of vectors orthogonal using gram Schmidt model. This model helps generate kth order polynomial and transfer model to linear, quadratic and cubic easily. This contrast method is always used for Ordinal Variables in which labels are equally spaced.

**Q6**. How can we convert a multi-class problem to a binary classiﬁcation problem when the objective is to understand the detractors among the group? Apply logistic regression on a binary classiﬁcation problem. Use all variables except the ones identiﬁed as leading to quasi-complete separation (use step-wise regression to build the model). Keep this in mind that you need to convert the attributes of survey questionnaires to ordinal variables before building the logistic model.

**Answer**:

There are many methods to perform transformation to binary from multiple class. Most prominent one’s are

1. Dummy Variable: This method employs transferring predictor class in n-1 columns and keeping values of class only in that column and set rest to zero. This helps training classifier for each binary class.
2. Reference Method: This method employs 1 to 1 and 1 vs all method of transformation of multiclass to binary class. We keep two binary class and rows to those classes and train classifiers on them.

We have used reference method one vs all to perform this binary classification. We have kept Detractors as **Positive class** and all the other are set to be “Not Detractors” or **negative class**

After converting all the survey response columns to ordinal factor form, we have run logistic model to find out variables which are causing quasi complete separation.

After removing those variables, we have deployed Step wise model to do feature selection and build model. Assuming that Detractors are positive class we need to correctly classify people as detractors and non-detractors. Cost of predicting false negative would not be beneficial as this will not increase NPS. We need to increase sensitivity of the model in order to decrease False Negative rate.

**Q7**: Compare the results of ensemble methods (Random Forest and Ada Boost) when applied to a multi-class classiﬁcation problem vis-a-vis a binary classiﬁcation problem.

We applied Random Forest and Ada boost to both sets of Binary and multi class data and found following as a result.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Accuracy | |  | Sensitivity | |
| Model | RF | Adaboost |  | RF | Adaboost |
|  |  |  |  |  |  |
| Normal |  |  |  |  |  |
| Multiclass | 67.86 | 67.2 |  | 47.73 | 47.73 |
| Binary | 90.11 | 90.65 |  | 40.9 | 43.18 |

**Assumption**: We have taken Detractors as positive class in order to identify them in real time. This will improve ability of MHE staff to predict Detractors and take appropriate actions to improve NPS score.

In order to improve NPS accurately we will need to reduce False Negative rate or increase sensitivity so that actual count of detractors can be identified.

After running Random Forest and Ada Boost on binary and multi class, we found that sensitivity is better in Adaboost and Random Forest and its around 47% for both cases. Since data is quite imbalanced, we need to deploy some resampling techniques to improve out result. Sensitivity score for Multiclass classification is better than Binary class.

Q8. Check the eﬀect of balancing methods (under-sampling, over-sampling, and SMOTE (Synthethic Minority Oversampling) on the performance of ensemble methods.

After running Random Forest and Ada boost on Normal data, we tried to transform imbalance in data via deploying techniques like Up sampling, SMOTE and down sampling. These techniques help remove imbalance in predictor class.

We ran Random Forest and Ada Boost for binary and Multiple class for each of the re sampling technique. And found results as follow.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Accuracy | |  | Sensitivity | |
| Model | RF | Adaboost |  | RF | Adaboost |
|  |  |  |  |  |  |
| Normal |  |  |  |  |  |
| Multiclass | 67.86 | 67.2 |  | 47.73 | 47.73 |
| Binary | 90.11 | 90.65 |  | 40.9 | 43.18 |
|  |  |  |  |  |  |
| Up sample |  |  |  |  |  |
| Multiclass | 66.48 | 62.36 |  | 50 | 54.55 |
| Binary | 90.66 | 85.16 |  | 40.91 | 61.36 |
|  |  |  |  |  |  |
| Down sample |  |  |  |  |  |
| Multiclass | 59.07 | 61.53 |  | 65.91 | 54.55 |
| Binary | 73.9 | 76.92 |  | 70.46 | 65.91 |
|  |  |  |  |  |  |
| SMOTE |  |  |  |  |  |
| Multiclass | 60.99 | 62.63 |  | 52.27 | 52.27 |
| Binary | 89.84 | 87.91 |  | 47.73 | 47.73 |

**Assumption**: We have taken Detractors as positive class in order to identify them in real time. This will improve ability of MHE staff to predict Detractors and take appropriate actions to improve NPS score.

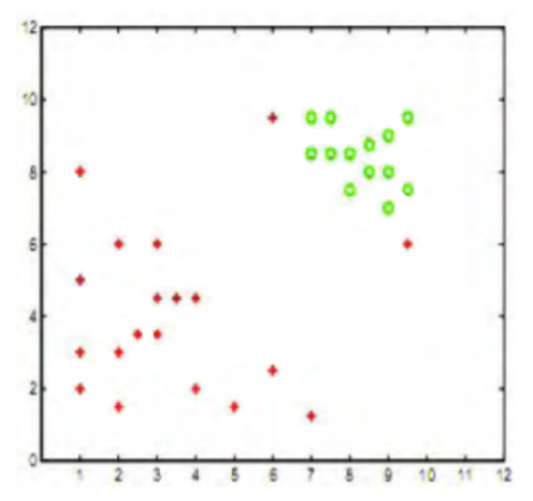
We can interpret following from above.

1. We observed that sensitivity for Random Forest multiclass was highest for Down sample technique and it reached to 66% from 47%.
2. For Binary class Random Forest, it reached to 70% from 40%. A good jump in sensitivity score which is most desired in our assumed case.
3. Though sensitivity has improved for each of the resampling technique, but Down Sampling has improved it Drastically. A huge increase in terms of 30 percentage points.

**Q8**. What should be the strategy for using the model to improve patient experience in the hospital and reduce proportion of detractors?

**Answers:** We have found some of the features which are directly impacting our predictor class of Detractors and Non-Detractors.

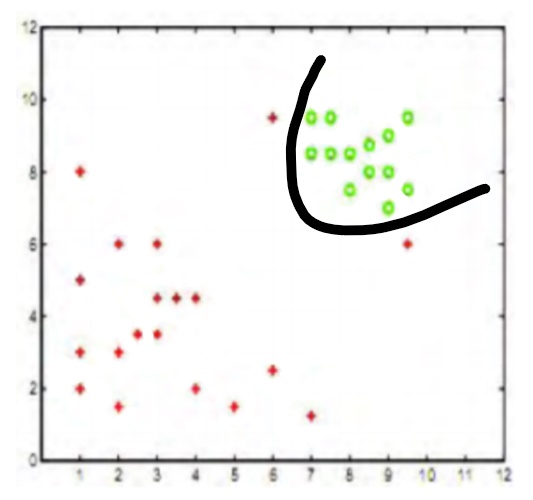
**Problem 2**. The original SVM proposed was a linear classier. As discussed in class, in order to make SVM non-linear we map the training data on to a higher dimensional feature space and then use a linear classier in the that space. This mapping can be done with the help of kernel functions. For this question assume that we are training an SVM with a quadratic kernel - i.e. our kernel function is a polynomial kernel of degree 2. This means the resulting decision boundary in the original feature space may be parabolic in nature. The dataset on which we are training is given below:



The slack penalty C will determine the location of the separating parabola. Please answer the following questions qualitatively.

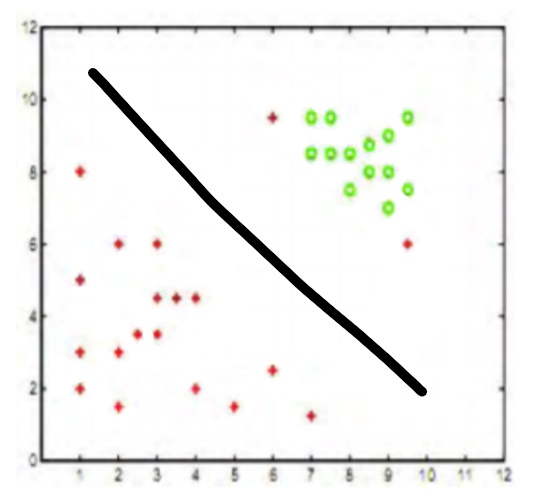
(a) Where would the decision boundary be for very large values of C? (Remember that we are using a quadratic kernel). Justify your answer in one sentence and then draw the decision boundary in the ﬁgure below.

**Answer**: As we are using high penalty value for C in above case so there would be any room for misclassification and every class will be classified correctly. Case will look like following.

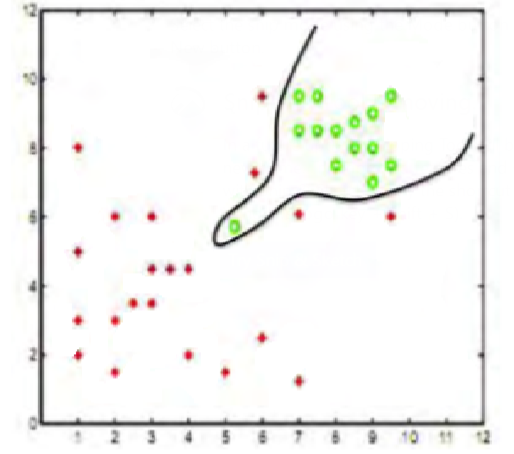


**(b)** Where would the decision boundary be for C nearly equal to 0? Justify your answer in one sentence and then draw the decision boundary in the ﬁgure below.

**Answer**: Value of C = 0 will impart zero penalty to misclassification and make coefficient of x2 zero. So classifier will be straight line classifying from between as follow.

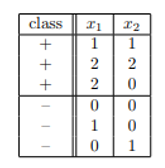


(c) Now suppose we add three more data points as shown in ﬁgure below. Now the data are not quadratically separable, therefore we decide to use a degree-5 kernel and ﬁnd the following decision boundary. Most probably, our SVM suﬀers from a phenomenon which will cause wrong classiﬁcation of new data points. Name that phenomenon, and in one sentence, explain what it is.



**Answer:** Excessive fitting to training data will lead to overfitting and less accuracy on test data.

**Problem 3**. Consider the following training data,

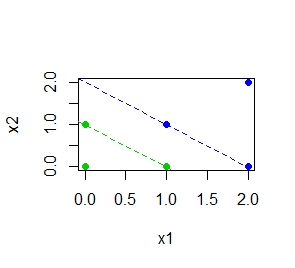


1. Plot these six training points. Are the classes {+, −} linearly separable?

**Answer:** after plotting these points, its much visible that these points are linearly separable.

**(b)** Construct the weight vector of the maximum margin hyperplane by inspection and identify the support vectors.

**Answer:** Required Image is as follow.



Points shown in Green and Blue are the support vectors.

1. If you remove one of the support vectors does the size of the optimal margin decrease, stay the same, or increase? Explain.

**Answer:** as it is much clear from picture that there are two support vectors on soft margin line. If we remove any one of them, one still remains on the line and position of margin does not changes.