

Class Prediction Error Pot: gives us ability to see how the portions of misclassification compared across different underlying true classification

Sensitivity = True Positives / (True Positives + False Negatives) = True Positive Rate = 90%

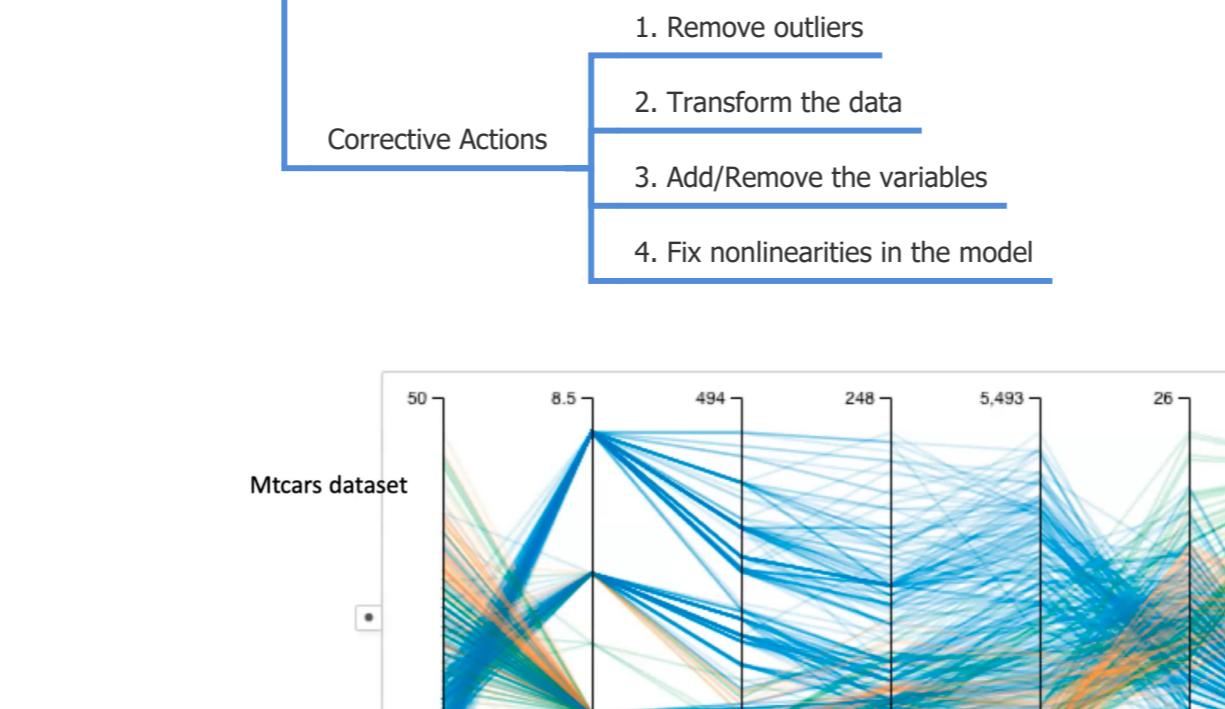
Tells us how well the model correctly identifies Positive cases

Specificity = True Negatives / (False Negatives + True Positives) = True Negative Rate = 80%

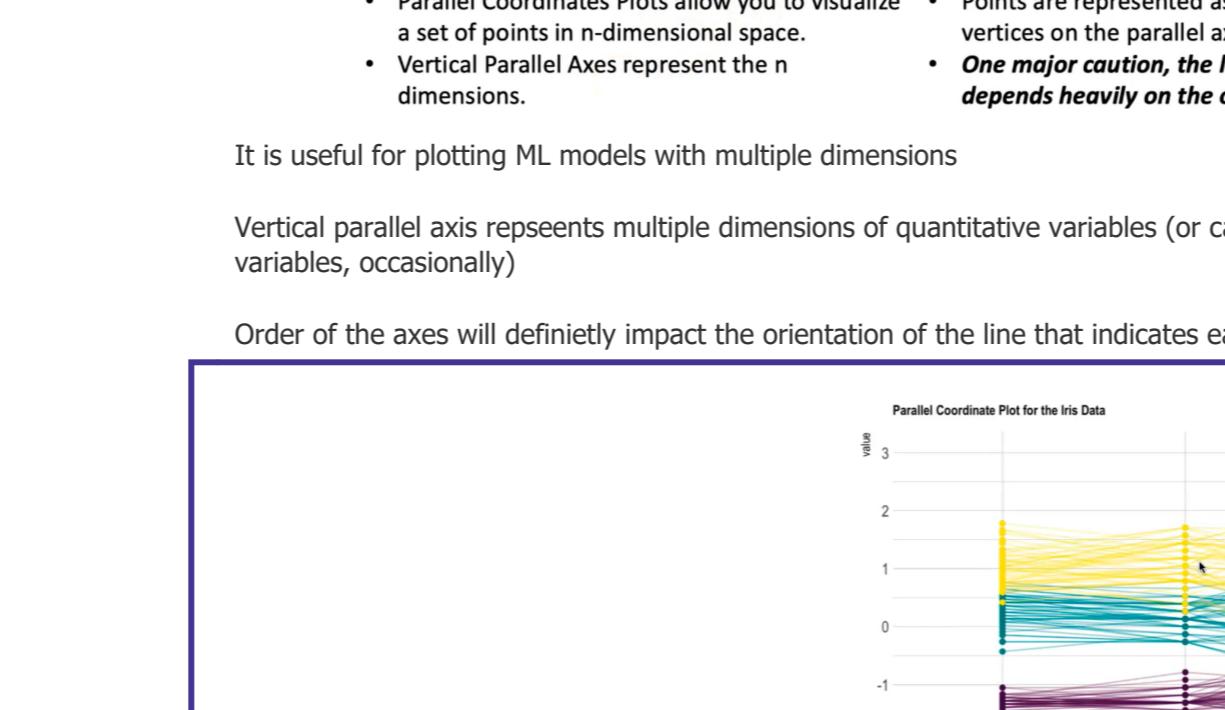
Tells us how well the model correctly identifies Negative cases

False Positive Rate = False Positives / (False Positives + True Negatives) = 1 - Specificity = 20%

Sensitivity and Specificity



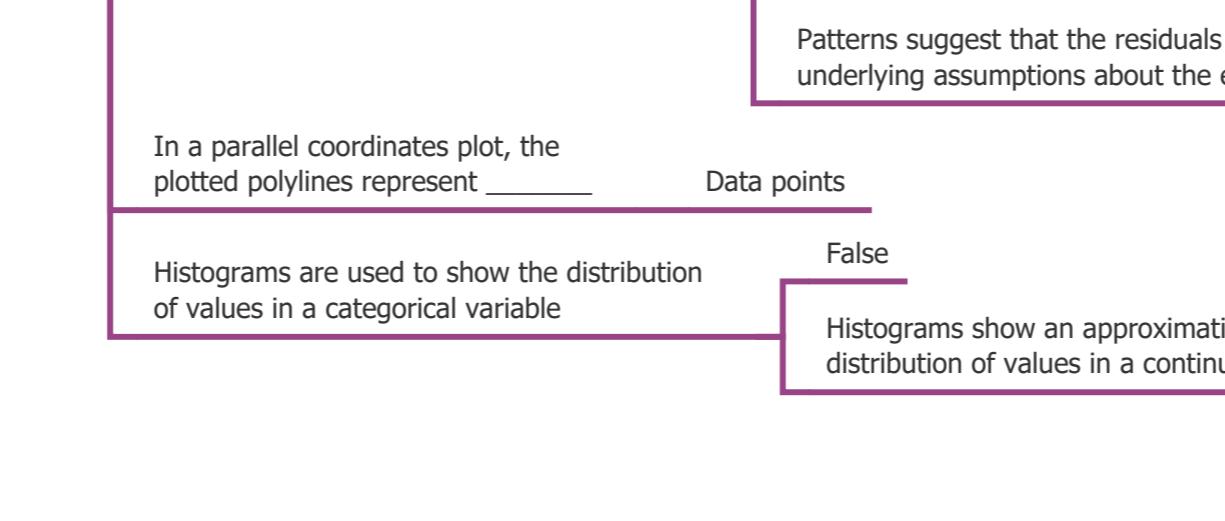
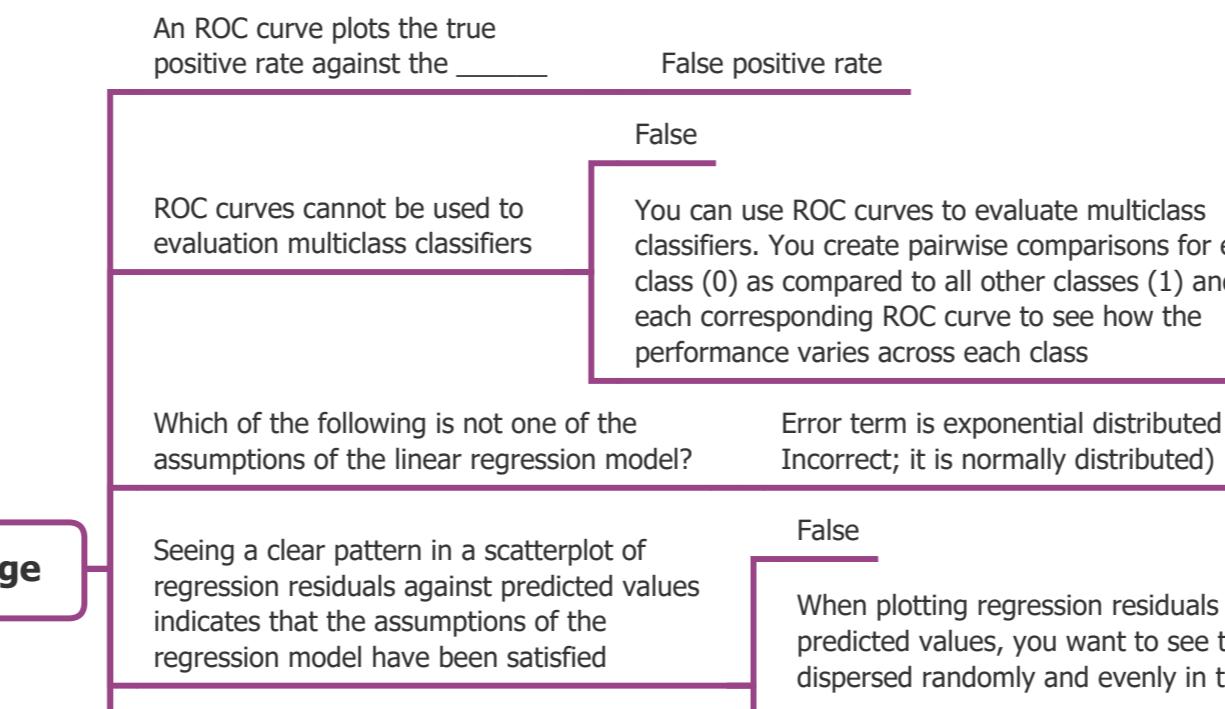
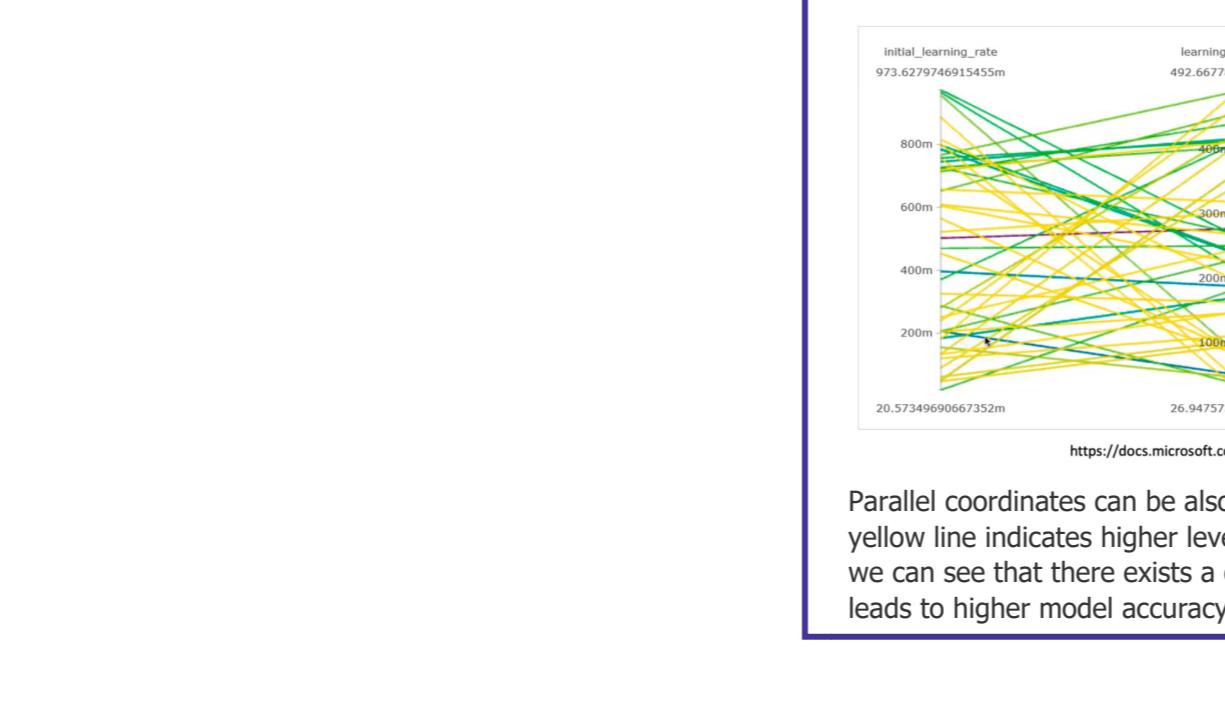
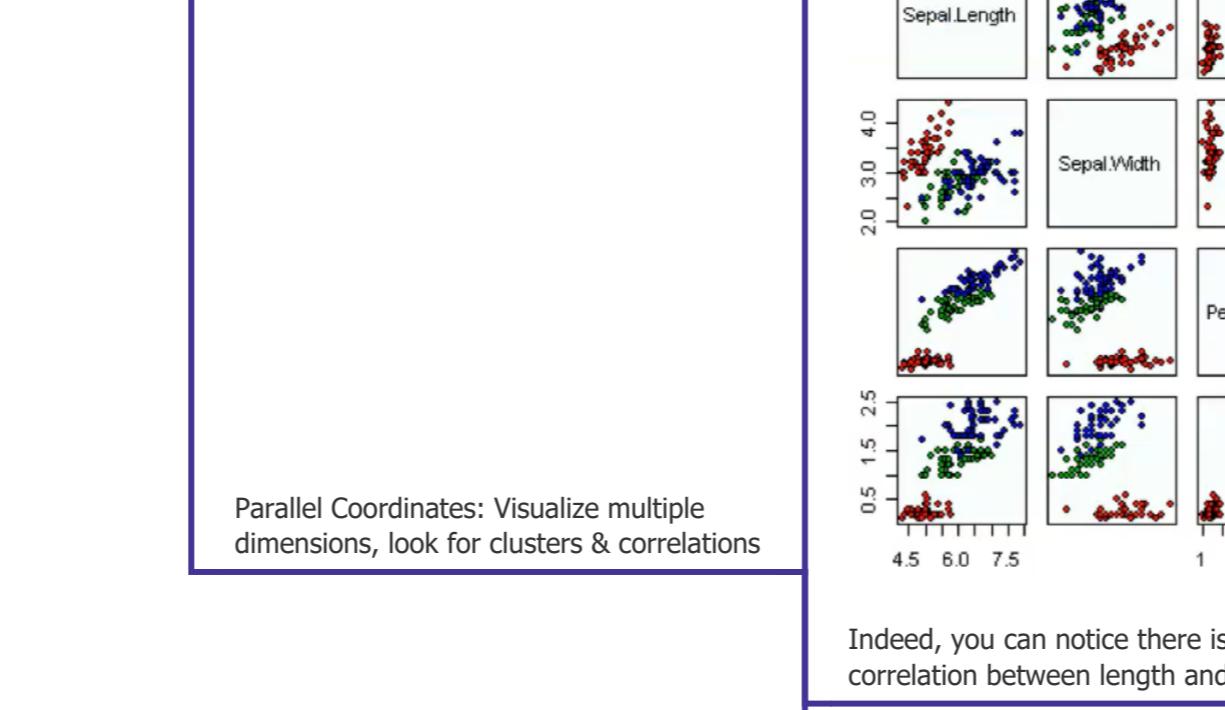
Receiver Operating Characteristics (ROC) Curves are graphical plots that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied



Area Under the Curve (AUC) is the probability that classifier will be able to distinguish between classes. Higher AUC is better!

AUC Benchmarks:

- 0.0-0.1: Poor
- 0.1-0.2: Bad
- 0.2-0.3: Fair
- 0.3-0.4: Acceptable/fair
- 0.4-0.5: Good
- 0.5-0.6: Very good
- 0.6-0.7: Excellent/good
- 0.7-0.8: Very excellent
- 0.8-0.9: Outstanding
- 0.9-1.0: Perfect



We want to ideally have 1 for TP and 0 for FP; model 4 would be the best one to use as it is the closest out of all to the ideal

Using AUC metric in ROC curve is useful way to compare different models in terms of their ability to distinguish the different classification



It performs multiple ROC analyses in a pairwise fashion - one class versus the rest; Look at the mean and individual class performance and compare across classifiers

For example, when you look at the ROC curve of class 0, it marks as TP when it correctly identifies class 0 as belonging in class 0 but FP when it incorrectly identifies class as belonging in class 1 or 2

ROC curves cannot be used to evaluate multiclass classifiers. You create pairwise comparisons for each class (0) as compared to all other classes (1) and plot the performance across each class

Error term is exponential distributed (Incorrect; it is normally distributed)

In the residual plots, we can look for patterns; if assumptions are met, residuals should look randomly and evenly dispersed around the 0 line



Heteroscedasticity is violating the assumption that variance is the same across all the residuals. As we move up in the size of the predicted value, we start to get larger and larger variance.

It is useful for plotting ML models with multiple dimensions

Vertical parallel axis represents multiple dimensions of quantitative variables (or categorical variables, occasionally)

Order of the axes will definitely impact the orientation of the line that indicates each observation



It is useful for plotting ML models with multiple dimensions

Vertical parallel axis represents multiple dimensions of quantitative variables (or categorical variables, occasionally)

Order of the axes will definitely impact the orientation of the line that indicates each observation



Edgar Anderson's Iris Data

Indeed, you can notice there is a high level of correlation between length and width

Parallel coordinates can also be used for hyper parameter tuning. In this instance, yellow lines represent the level of accuracy in terms of model building procedure, and we can see that there even a cluster of yellow lines, indicating that lower batch size leads to higher model accuracy



ROC curves cannot be used to evaluate multiclass classifiers

You can use ROC curves to evaluate multiclass classifiers. You create pairwise comparisons for each class (0) as compared to all other classes (1) and plot the performance across each class

Which of the following is not one of the assumptions of the linear regression model?

Error term has mean of 0

Error term has constant variance

Error term are independent

Errors are normally distributed

Seeing a clear pattern in a scatterplot of regression residuals against predicted values indicates that assumptions of the regression model have been satisfied

When plotting regression residuals against predicted values, you want to see the data points dispersed randomly and evenly in the graph

Patterns suggest that the residuals are violating underlying assumptions about the error term



Notice how other points are either clustered or one or couple points are placed away from the rest of the points

Corrective Actions:

1. Remove outliers
2. Transform the data
3. Add/Remove the variables
4. Fix nonlinearities in the model

There are some values that have really large values of x; if we flipped the axis, then some y values would be unbalanced



It is useful for plotting ML models with multiple dimensions

Vertical parallel axis represents multiple dimensions of quantitative variables (or categorical variables, occasionally)

Order of the axes will definitely impact the orientation of the line that indicates each observation



It is useful for plotting ML models with multiple dimensions

Vertical parallel axis represents multiple dimensions of quantitative variables (or categorical variables, occasionally)

Order of the axes will definitely impact the orientation of the line that indicates each observation



Edgar Anderson's Iris Data

Indeed, you can notice there is a high level of correlation between length and width

Parallel coordinates can also be used for hyper parameter tuning. In this instance, yellow lines represent the level of accuracy in terms of model building procedure, and we can see that there even a cluster of yellow lines, indicating that lower batch size leads to higher model accuracy



ROC curves cannot be used to evaluate multiclass classifiers

You can use ROC curves to evaluate multiclass classifiers. You create pairwise comparisons for each class (0) as compared to all other classes (1) and plot the performance across each class

Which of the following is not one of the assumptions of the linear regression model?

Error term has mean of 0

Error term has constant variance

Error term are independent

Errors are normally distributed

Seeing a clear pattern in a scatterplot of regression residuals against predicted values indicates that assumptions of the regression model have been satisfied

When plotting regression residuals against predicted values, you want to see the data points dispersed randomly and evenly in the graph

Patterns suggest that the residuals are violating underlying assumptions about the error term



Notice how other points are either clustered or one or couple points are placed away from the rest of the points

Corrective Actions:

1. Remove outliers
2. Transform the data
3. Add/Remove the variables
4. Fix nonlinearities in the model

There are some values that have really large values of x; if we flipped the axis, then some y values would be unbalanced



It is useful for plotting ML models with multiple dimensions

Vertical parallel axis represents multiple dimensions of quantitative variables (or categorical variables, occasionally)

Order of the axes will definitely impact the orientation of the line that indicates each observation



It is useful for plotting ML models with multiple dimensions

Vertical parallel axis represents multiple dimensions of quantitative variables (or categorical variables, occasionally)

Order of the axes will definitely impact the orientation of the line that indicates each observation



Edgar Anderson's Iris Data

Indeed, you can notice there is a high level of correlation between length and width

Parallel coordinates can also be used for hyper parameter tuning. In this instance, yellow lines represent the level of accuracy in terms of model building procedure, and we can see that there even a cluster of yellow lines, indicating that lower batch size leads to higher model accuracy



ROC curves cannot be used to evaluate multiclass classifiers

You can use ROC curves to evaluate multiclass classifiers. You create pairwise comparisons for each class (0) as compared to all other classes (1) and plot the performance across each class

Which of the following is not one of the assumptions of the linear regression model?

Error term has mean of 0

Error term has constant variance

Error term are independent

Errors are normally distributed

Seeing a clear pattern in a scatterplot of regression residuals against predicted values indicates that assumptions of the regression model have been satisfied

When plotting regression residuals against predicted values, you want to see the data points dispersed randomly and evenly in the graph

Patterns suggest that the residuals are violating underlying assumptions about the error term



Notice how other points are either clustered or one or couple points are placed away from the rest of the points

Corrective Actions:

1. Remove outliers
2. Transform the data
3. Add/Remove the variables
4. Fix nonlinearities in the model

There are some values that have really large values of x; if we flipped the axis, then some y values would be unbalanced



It is useful for plotting ML models with multiple dimensions

Vertical parallel axis represents multiple dimensions of quantitative variables (or categorical variables, occasionally)

Order of the axes will definitely impact the orientation of the line that indicates each observation



It is useful for plotting ML models with multiple dimensions

Vertical parallel axis represents multiple dimensions of quantitative variables (or categorical variables, occasionally)

Order of the axes will definitely impact the orientation of the line that indicates each observation

