**Regularization** normally tries to reduce or penalize the complexity of the model. Regularization techniques applied with logistic regression mostly tend to penalize large coefficients 𝑏₀, 𝑏₁, …, 𝑏ᵣ:

* **L1 regularization** penalizes the LLF with the scaled sum of the absolute values of the weights: |𝑏₀|+|𝑏₁|+⋯+|𝑏ᵣ|.
* **L2 regularization** penalizes the LLF with the scaled sum of the squares of the weights: 𝑏₀²+𝑏₁²+⋯+𝑏ᵣ².

As LogisticRegression applies an L2 regularization by default, the result

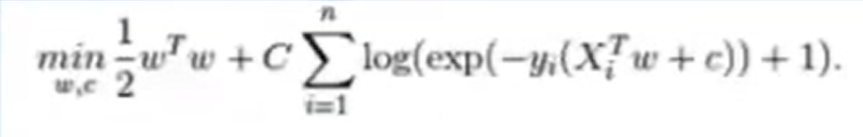
looks similar to Ridge. Stronger regularization pushes coefficients more and more towards zero, though coefficients never become exactly zero.

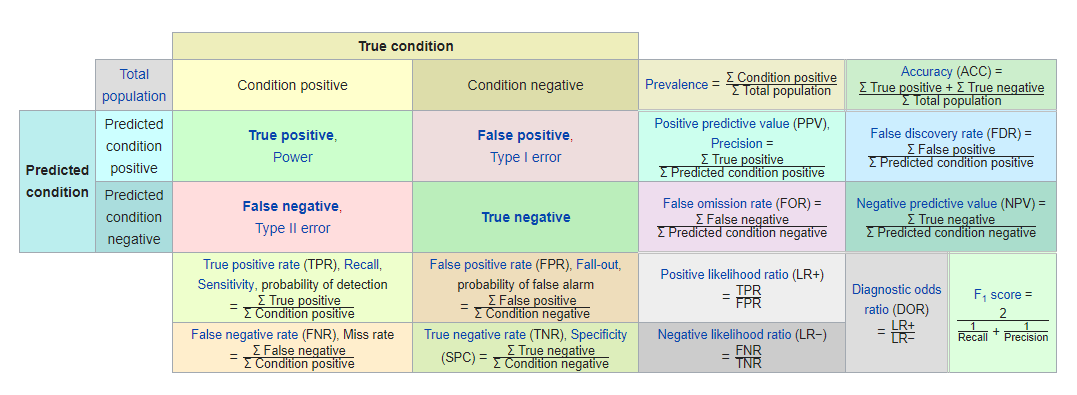
If we desire a more interpretable model, using L1 regularization might help

LASSO Regularisation (l1)



Ridge Regularisation( l2)





**True Positive Rate (TPR)** - It indicates how many positive values, out of all the positive values, have been correctly predicted. The formula to calculate the true positive rate is (TP/TP + FN). Also, TPR = 1 - False Negative Rate. It is also known as Sensitivity or Recall.

**False Positive Rate (FPR)** - It indicates how many negative values, out of all the negative values, have been incorrectly predicted. The formula to calculate the false positive rate is (FP/FP + TN). Also, FPR = 1 - True Negative Rate.

**True Negative Rate (TNR)** - It indicates how many negative values, out of all the negative values, have been correctly predicted. The formula to calculate the true negative rate is (TN/TN + FP). It is also known as Specificity.

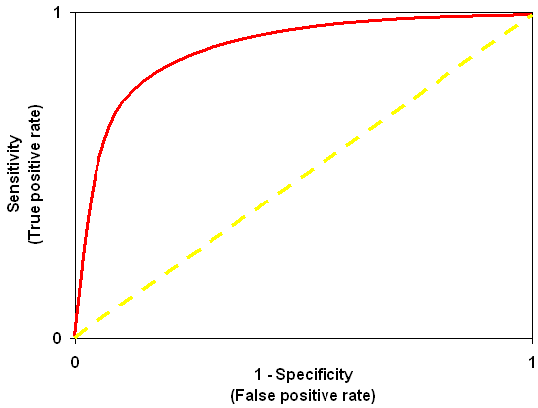
**False Negative Rate (FNR)** - It indicates how many positive values, out of all the positive values, have been incorrectly predicted. The formula to calculate false negative rate is (FN/FN + TP).

**Precision:** It indicates how many values, out of all the predicted positive values, are actually positive. It is formulated as:(TP / TP + FP).

**Recall**: It indicates how many values, out of all the Actual positive values, how many are predicted as positive. It is formulated as:(TP / TP + FN). It is also referred as Sensitivity

**F Score**: F score is the harmonic mean of precision and recall. It lies between 0 and 1. Higher the value, better the model. It is formulated as 2((precision\*recall) / (precision+recall)).

**Receiver Operator Characteristic (ROC):** ROC determines the accuracy of a classification model at a user defined threshold value. It determines the model's accuracy using Area Under Curve (AUC). Higher the area, better the model. ROC is plotted between True Positive Rate (Y axis) and False Positive Rate (X Axis). The yellow line represents the ROC curve at 0.5 threshold.



1. **Akaike Information Criteria (AIC):** You can look at AIC as counterpart of adjusted r square in multiple regression. It's an important indicator of model fit. It follows the rule: **Smaller the better**. AIC penalizes increasing number of coefficients in the model.

The [Akaike Information Criterion](https://en.wikipedia.org/wiki/Akaike_information_criterion), or AIC for short, is a method for scoring and selecting a model.

*To use AIC for model selection, we simply choose the model giving smallest AIC over the set of models considered.*

* AIC = -2/N \* LL + 2 \* k/N

Where *N* is the number of examples in the training dataset, *LL* is the log-likelihood of the model on the training dataset, and *k* is the number of parameters in the model.

the AIC statistic penalizes complex models less, meaning that it may put more emphasis on model performance on the training dataset, and, in turn, select more complex models.

# calculate aic for regression

def calculate\_aic(n, mse, num\_params):

aic = n \* log(mse) + 2 \* num\_params

return aic

1. **Bayesian Information Criteria (BIC)**

The [Bayesian Information Criterion](https://en.wikipedia.org/wiki/Bayesian_information_criterion), or BIC for short, is a method for scoring and selecting a model.

* BIC = -2 \* LL + log(N) \* k

Where *log()* has the base-e called the natural logarithm, *LL* is the log-likelihood of the model, *N* is the number of examples in the training dataset, and *k* is the number of parameters in the model.

The score as defined above is minimized, e.g. the model with the lowest BIC is selected