

**WEB API**

Amigo Ricardo: por favor describe el proceso para poder acceder al API de Tweeter

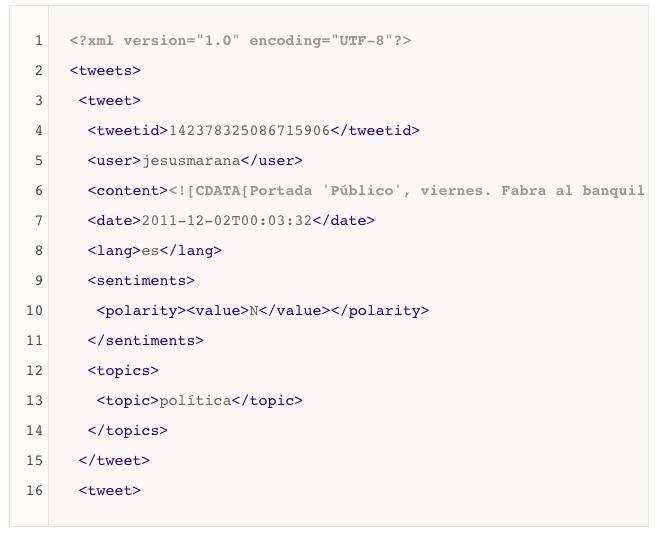
**SOCIAL MEDIA PLATFORM**

* Downloading the tweets (Tocaya+Pandis hicieron esto)

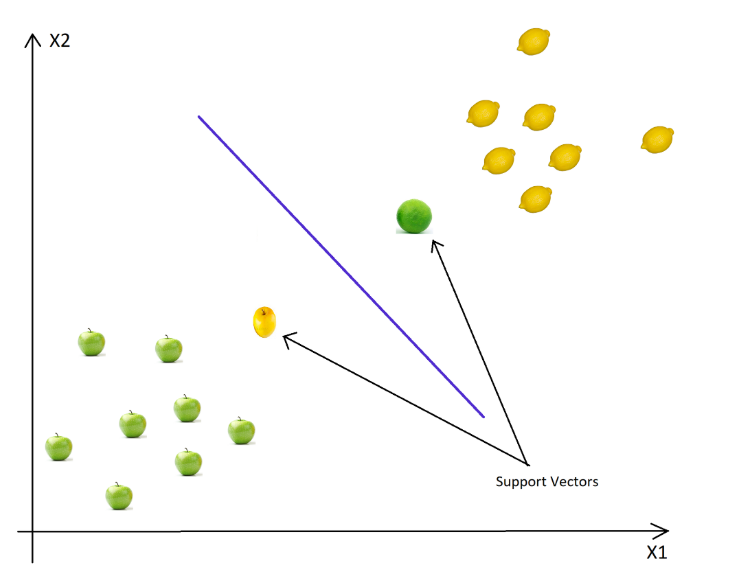
**ANALYTICS ENGINE**

Steps for generating the algorithm for detecting tweet’s sentiment:

1. In Python, gather and parse pre-tagged tweets from Sentiment Analysis in Spanish Workshop (different topics: politics, TV, etc). Format of the datasets:

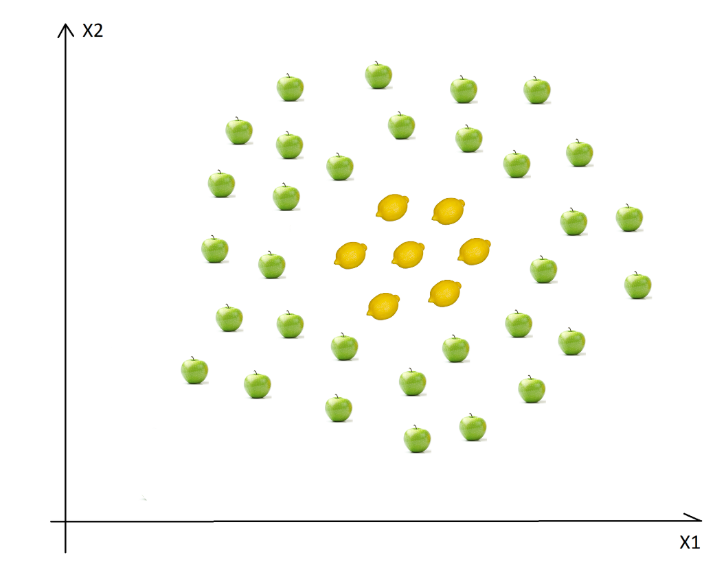


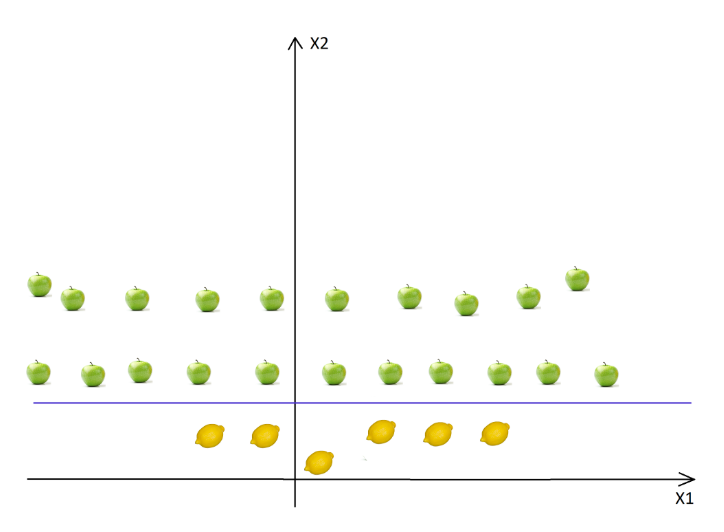
1. Binarize the sentiment: tweets are classified by level of sentiment (P+, P, NEU, N, N+) so we binarized it such that negative=0 and positive=1
2. We processed the text before creating the algorithm:
   1. Turned the text into a matrix of token counts with scikit learn’s *CountVectorizer* in order to work with those vectors instead of the raw text
   2. Tokenized the text
   3. Turn all words to lowercase
   4. Removed stopwords
   5. Stemmed the words with *SnowballStemmer*, a NLT that supports Spanish: kept the stem (e.g., stem of beautifully and beautiful is beauti)
3. Choose a ML classifier: Linear Support Vector Classifier is known to perform well in binary classification problems



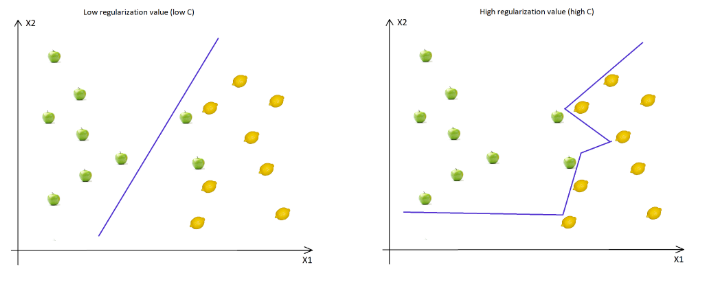
1. Tuning parameters with GridSearchCV:

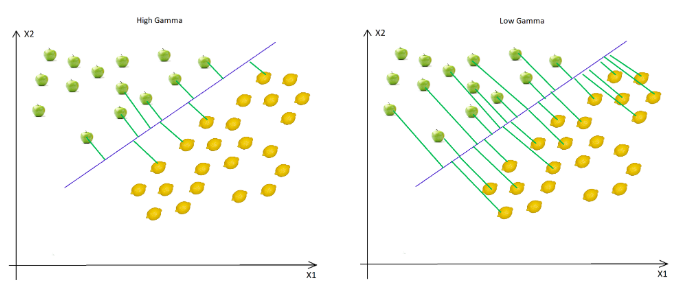
kernel : transforms the problem in order to be able to use linear algebra



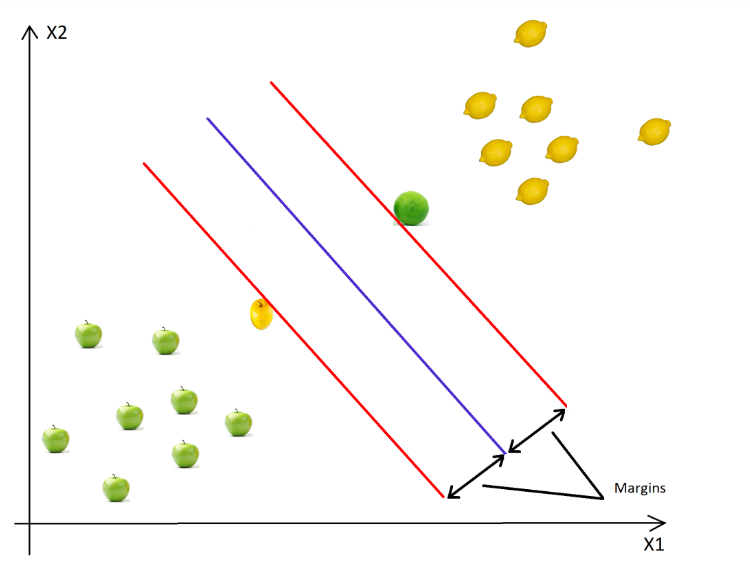


Regularization: defines where the separation is larger for both classes



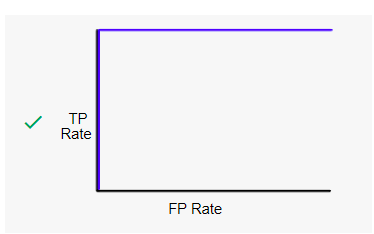
Gamma: defines how far are the farthest points considered 

Margin: defines where the separation is the larger for both classes

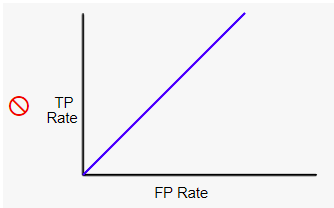


1. Model evaluation: AUC

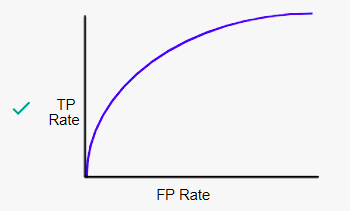
Best possible: AUC = 1 predicts true positives always



AUC = 0.5 🡪 predicts true positives only 50% of the times



AUC = 0.79 (our model) 🡪 predicts true positives most of the times



1. Polarity prediction: once we have our model we predict the polarity of our Metro CDMX tweets. We decided to name our model “Sheldon” as, even when he does a great job most of the times









He struggles with sarcasm…



**DATA VIZ AND REPORTS 🡪 Ricardo y León**