Рассмотрим задачу классификации сообщения на спам и на реальные письма.

Придумайте вероятностную модель превращающую письмо в разряженный вектор признаков. Модель должна поддерживать n-граммы и учитывать как заголовок, так и содержание письма.

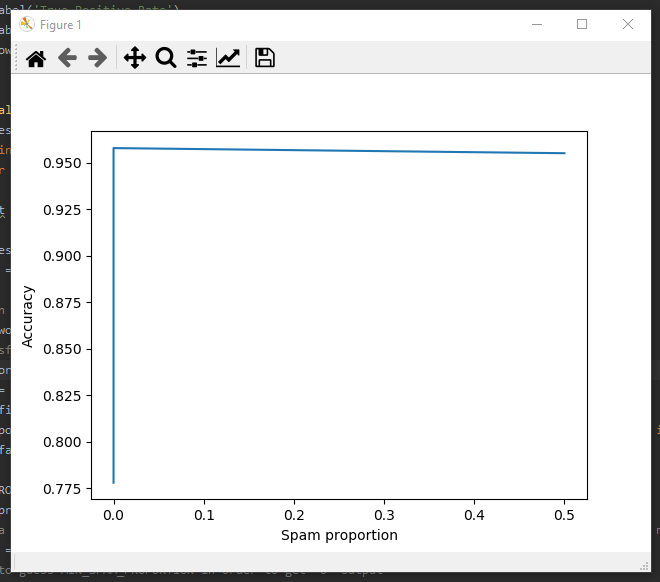
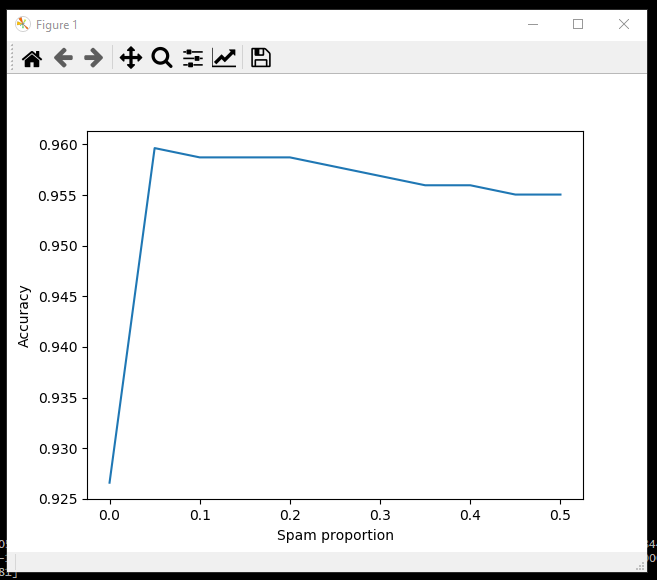
Постройте ROC кривую для выбранной модели. Посчитайте точность, используя перекрестную проверку.

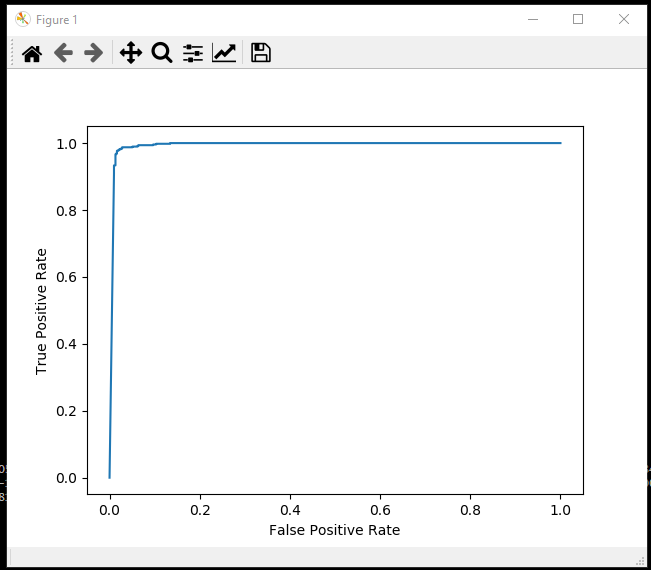
Контролируя априорное распределение добейтесь того, чтобы ни одно реальное сообщение не было классифицировано как спам. Постройте график зависимости точности от выбранного параметра λ, где λ изменяется от значения по умолчанию (λspam = λlegit), до найденного значения в предыдущем пункте.

Общая логика – некоторые слова часто встречаются в спаме, можно посчитать число и вероятность их вхождений. Тогда вероятность, что письмо – спам – произведение вероятностей, что слово спамовое \* априорная вероятность получить спамовое письмо (по теореме Байеса). А классификатор наивный, потому что мы полагаем, что слова в письме между собой не связаны (но для грубой оценки пойдет).

Подобрана минимальная вероятность спама в 10 в минус 85 степени (почти ноль), при которой ни одно реальное письмо не классифицируется как спам. Модель обучается, дает предсказания, которые проверяются на ложноположительность.

По полученному минимальному значению вероятности строю график зависимости точности от априорной вероятности получения спама. Так как полученная величина вероятности получения спама не очень соответствует действительности, то и точность на отрезке малых вероятностей неприлично мала. Оптимальное значение 0.9596330275229358 достигается при величине вероятности спама порядка 0.05. Графики: на первом с шагом 0.5, на втором хорошо просчитаны мелкие вероятности вплоть до 10 в -85 степени (шаг прогрессивный, 10 в -5 степени) – на малых значениях скачков нет.





ROC-кривая: TPR полностью совпадает с полнотой, и показывает долю верно предсказанных классов у объектов, относящихся к положительному классу. FPR — это доля неправильно предсказанных классов среди объектов отрицательного класса.

Так как график выше прямой у = х, классификатор работает успешно