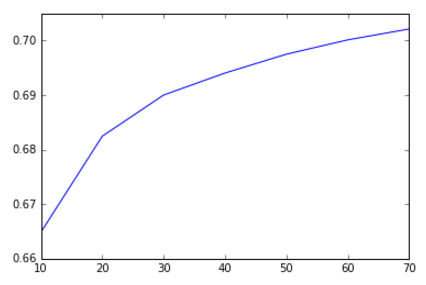
**Часть 1:**

1. Список признаков, для которых имеются пропуски:

['first\_blood\_time', 'first\_blood\_team', 'first\_blood\_player1', 'first\_blood\_player2', 'radiant\_bottle\_time', 'radiant\_courier\_time', 'radiant\_flying\_courier\_time', 'radiant\_first\_ward\_time', 'dire\_bottle\_time', 'dire\_courier\_time', 'dire\_flying\_courier\_time', 'dire\_first\_ward\_time']

Это признаки-события, которые могли не произойти к 5-ой минуте матча. То есть т.к. все эти колонки даны на момент 5-ой минуты, то к этому времени могли ещё не убить ни одного героя или не найти артефакты/ предметы.

1. Столбец называется «radiant\_win»
2. Качество 0,69. Для оценки потребовалось 2 мин 5 секунд
3. Имеет смысл использовать больше 30 деревьев, потому что мы видим улучшение качества при увеличении количество деревьев до 40, 50 и далее.



Для ускорения процесса работы программы можно менять параметры метода обучения, например: уменьшить глубину деревьев (max\_depth), min\_samples\_leaf, min\_samples\_split, max\_features, etc. Также можно проводить поиск оптимального параметра на кросс-валидации только на части обучающей выборки. Ещё можно уменьшить количество используемых факторов за счёт выброса незначимых факторов. Или изменить сам метод обучения

**Часть 2:**

1. Максимальное качество на кросс валидации 0.716649. Оно достигается при C = 0.004. Качество на кросс валидации на 0.02 выше, чем у градиентного бустинга (при 30 деревьях). Связано это с тем, что мы недообучили градиентный бустинг. Скорее всего при большем количестве деревьев качество модели градиентого бустинга значительно улучшится. Логистическая регрессия работает значительно быстрее, чем расчёт градиентного бустинга.
2. Новое качество модели 0,71666, которое достигается при С = 0.005. То есть качество модели немного улучшилось. Скорее всего это связано с тем, что при большем количестве факторов модель была переобучена.
3. 108 идентификаторов героев
4. Новое качество 0.71715577 (достигается при C = 0.005). Качество немного улучшилось по сравнению с предыдущими подпунктами. Улучшение качества связано с добавлением новых переменных.
5. Минимальная вероятность - 0.00738, максимальная - 0.987311