1. Расскажите, как работает регуляризация в решающих деревьях, какие параметры мы штрафуем в данных алгоритмах?

Чтобы избежать переобучения, мы должны добавить параметры регуляризации в модель.  
Начнем с параметра максимальной глубины – max\_depth Параметр max\_depth хорошо справляется с регуляризацией модели, и мы уже не так сильно переобучаемся. Доля верных ответов возростает..

Еще важный параметр min\_samples\_leaf, он также выполняет функцию регуляризатора.

Еще параметр max\_features. Для задач классификации по умолчанию используется n, где n — число признаков.

1. По какому принципу рассчитывается "важность признака (feature\_importance)" в ансамблях деревьев?

Обычный способ вычисления значений важности признаков одного дерева заключается в следующем:

1. вы инициализируете массив feature\_importances всех нулей размером n\_features .
2. вы проходите по дереву: для каждого внутреннего узла, который разбивается на функцию i , вы вычисляете уменьшение ошибки этого узла, умноженное на количество выборок, которые были направлены в узел, и добавляете это количество к feature\_importances .

Уменьшение ошибки зависит от используемого критерия примеси (например, Джини, энтропия, MSE, ...).-это примесь набора примеров, которая направляется во внутренний узел за вычетом суммы примесей двух разделов, созданных расщеплением.

Важно, чтобы эти значения относились к определенному набору данных (как уменьшение ошибок, так и количество выборок являются специфичными для набора данных), поэтому эти значения не могут быть сопоставлены между различными наборами данных.

Насколько я знаю, существуют альтернативные способы вычисления значений важности признаков в деревьях принятия решений. Краткое описание вышеприведенного метода можно найти в книге Тревора Хасти, Роберта Тибсирани и Джерома Фридмана.