

1. Что такое L2 регуляризация, в чем отличие от L1?

L2 регуляризация предотвращает переобучение модели путём запрета на несоразмерно большие весовые коэффициенты. При L2 регуляризации дополнительный член является квадратичной функцией (при L1 регуляризации – модулем).

$$L(x, y) \equiv \sum_{i=1}^n (y_i - h_{\theta}(x_i))^2 + \lambda \sum_{i=1}^n \theta_i^2,$$

где λ является гиперпараметром, который контролирует интенсивность штрафа.

Когда $\lambda \rightarrow 0$, результаты аналогичны линейной регрессии.

Когда $\lambda \rightarrow \infty$, все особенности уменьшены до 0.

Отличия:

- L1 штрафует сумму абсолютных значений весов, а L2 штрафует сумму квадратичных весов.
- L1 имеет разреженное решение, L2 имеет не разреженное решение.
- L1 имеет несколько решений, у L2 есть одно решение.
- L1 имеет встроенный выбор функций, L2 не имеет выбора функций.
- L1 устойчив к выбросам, L2 не устойчив к выбросам.
- L1 генерирует модели, которые просты и понятны, но не могут выучить сложные шаблоны, а L2 дает лучший прогноз, когда выходная переменная является функцией всех входных функций.

2. Для чего используется LambdaCallback?

LambdaCallback используется для создания простых пользовательских обратных вызовов на лету. Этот callback строится с помощью анонимных функций. За счет этого обратная связь будет быстрее.

3. Почему выбраны именно такие модели для ансамбля?

Для обработки текстовых данных часто используют рекуррентные сети, потому была выбрана оптимальная модель такой сети. Многослойный перцептрон дал результаты хуже, от него отказалась. Сверточная сеть давала результаты даже лучше, чем рекуррентная сеть, потому она была использована. Была использована сверточную рекуррентную сеть для того, чтобы полученная сеть выявляла другие закономерности.