

1. Что такое многозадачное обучение?

Многозадачное обучение – способ улучшить обобщаемость путем пулинга примеров, возникающих в нескольких задачах. Форма многозадачного обучения предусматривает, что разные задачи обучения с учителем (предсказание y_i при известном x) разделяют один и тот же вход x , а также некоторое промежуточное представление, запоминающее общий пул факторов. В общем случае модель можно разделить на две части:

a) Параметры, специфичные для конкретной задачи (для достижения хорошей обобщаемости им нужны только примеры, относящиеся к «своей» задаче).

b) Общие параметры, разделяемые всеми задачами (они извлекают выгоду из организации пула данных всех задач).

Улучшенной обобщаемости и ограничения ошибки обобщения можно достичь благодаря разделяемым параметрам, статистическую мощность которых можно значительно повысить. Для этого нужно, чтобы выполнялись предположения относительно статистической связи между различными задачами, выражающие тот факт, что у некоторых задач действительно имеется что-то общее. Иначе говоря, среди факторов, объясняющих вариативность, наблюдаемую в данных, ассоциированных с разными задачами, некоторые являются общими для двух или более задач.

2. 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).

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

λ в уравнении 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 квадратных 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.

3. В отчете не увидел ничего про ансамблирование. Как Вы его выполняли?

Сначала высчитывается предикт для обеих моделей. Если их среднее арифметическое больше 0.5, значит выход 1, иначе 0. Ансамблирование достигается за счет учета предиктов выше описанным способом.

```
pred1 = model.predict(input)
pred2 = model_2.predict(input)
pred = [1 if (pred1[i] + pred2[i]) / 2 > 0.5 else 0 for i in range(len(pred1))]
return np.array(pred)
```

4. В чем разница между GRU и LSTM юнитами?

В большинстве случаев GRU работают так же, как LSTM. Основное отличие GRU от LSTM заключается в том, что один вентиляционный блок одновременно управляет и коэффициентом забывания, и решением об обновлении блока состояния. Различие между GRU и LSTM также в том, что GRU быстрее и проще в эксплуатации, однако обладает немного меньшими возможностями.