

Часть 7: Регуляризация нейронных сетей

...

Романов Михаил, Игорь Слинько

В чем проблема?

Рассмотрим обратную задачу:

$$Ax = b$$

В чем проблема?

Рассмотрим обратную задачу:

$$Ax = b$$

$$\|Ax - b\|_2^2 \rightarrow \min$$

В чем проблема?

Рассмотрим обратную задачу:

$$Ax = b$$

$$\|Ax - b\|_2^2 \rightarrow \min_x$$

$$\text{let } A = \begin{bmatrix} \end{bmatrix}$$

В чем проблема?

Рассмотрим обратную задачу:

$$Ax = b$$

$$\|Ax - b\|_2^2 \rightarrow \min_x$$

$$\text{let } A = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \Rightarrow \exists! x$$

$$\text{let } A = \begin{bmatrix} 1 & 1 \end{bmatrix}$$

В чем проблема?

Рассмотрим обратную задачу:

$$Ax = b$$

$$\|Ax - b\|_2^2 \rightarrow \min_x$$

$$\text{let } A = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \Rightarrow \exists! x$$

$$\text{let } A = \begin{bmatrix} 1 & 1 \end{bmatrix} \quad \# x \rightarrow \infty$$

$$Ax_0 = 0$$

$$Ax_y = b$$

$$A(x_y + x_0) = b$$

В чем проблема?

Рассмотрим обратную задачу:

$$Ax = b$$

$$\|Ax - b\|_2^2 \rightarrow \min_x$$

$$\text{let } A = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \Rightarrow \exists! x$$

$$\text{let } A = \begin{bmatrix} 1 & 1 \end{bmatrix} \quad \# x \rightarrow \infty$$

$$Ax_0 = 0$$

$$Ax_y = b$$

$$A(x_y + x_0) = b$$

Однако не все x одинаково полезны
(бритва Оккама: при прочих равных
лучше то, что проще):

$$\|Ax - b\|_2^2 + \lambda \|x\|_2^2 \rightarrow \min_x$$

Статистическое обоснование регуляризации

$$p(x|b) =$$

Статистическое обоснование регуляризации

$$p(x|b) = \frac{p(b|x) \cdot p(x)}{p(b)} \rightarrow \max_x$$

Статистическое обоснование регуляризации

$$p(x|b) = \frac{p(b|x) \cdot p(x)}{p(b)} \rightarrow \max_x$$

$$\log p(b|x) + \log p(x) - \log p(b) \rightarrow \max_x$$

Статистическое обоснование регуляризации

$$p(x|b) = \frac{p(b|x) \cdot p(x)}{p(b)} \rightarrow \max_x$$

$$\log p(b|x) + \log p(x) - \log p(b) \rightarrow \max_x$$

$$\log p(b|x) + \log p(x) \rightarrow \max_x$$

Статистическое обоснование регуляризации

$$\log p(b|x) + \log p(x) \rightarrow \max_x$$

$$p(b|x) = d \cdot \exp\left(-\frac{\|F(x) - b\|_2^2}{2\sigma^2}\right)$$

$$p(x) = d' \cdot \exp(-\|x\|_2^2 \cdot \lambda)$$

Регуляризация и нейронные сети

- Когда #параметров > #данных
- Или когда модель имеет склонность к переобучению (похожие данные и т.д.)

$$\mathcal{L}(f(X, w), Y) \rightarrow \min_w$$

Регуляризация и нейронные сети

- Когда #параметров > #данных
- Или когда модель имеет склонность к переобучению (похожие данные и т.д.)

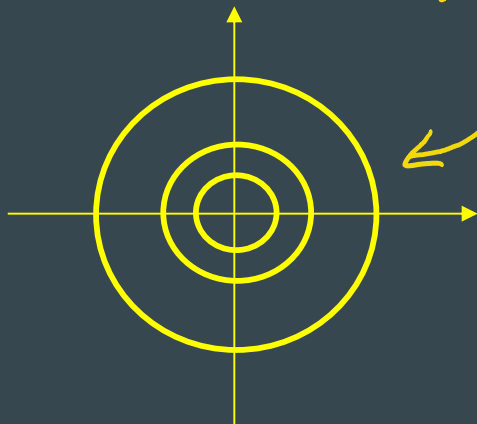
$$\mathcal{L}(f(X, w), Y) \rightarrow \min_w$$

$$\mathcal{L}(f(x, w), Y) + \lambda \log p(w) \rightarrow \min_w$$

Регуляризации и их свойства

L2

$$p(w) = \alpha \cdot \exp\left(-\sum_i w_i^2\right) \Leftrightarrow + \lambda \|w\|_2^2$$

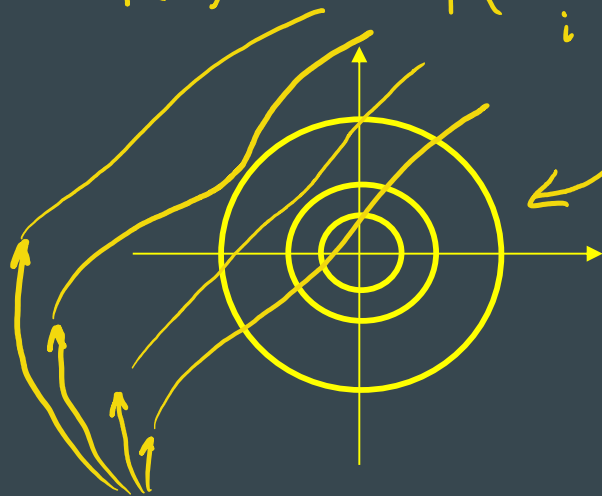


Линии уровня регуляризации

Регуляризации и их свойства

L_2

$$p(w) = \alpha \cdot \exp\left(-\sum_i w_i^2\right) \Leftrightarrow + \lambda \|w\|_2^2$$



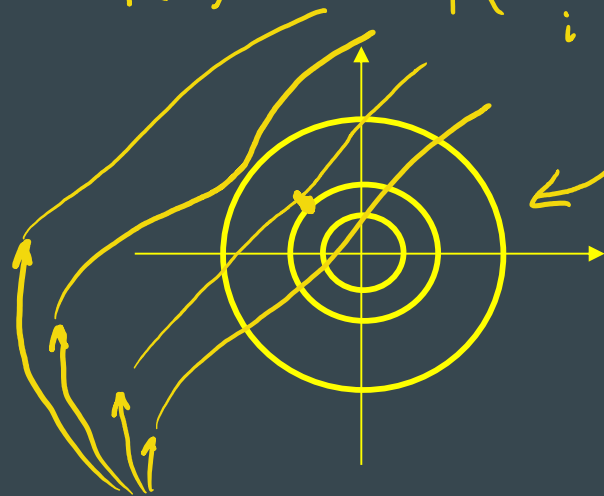
Линии уровня регуляризации

Линии уровня
функции потерь

Регуляризации и их свойства

L_2

$$p(w) = \alpha \cdot \exp\left(-\sum_i w_i^2\right) \Leftrightarrow + \lambda \|w\|_2^2$$



Минимальная линия уровня

Линии уровня регуляризации

Линии уровня
функции потерь

Регуляризации и их свойства

L_2

$$p(w) = \alpha \cdot \exp\left(-\sum_i w_i^2\right) \Leftrightarrow + \lambda \|w\|_2^2$$

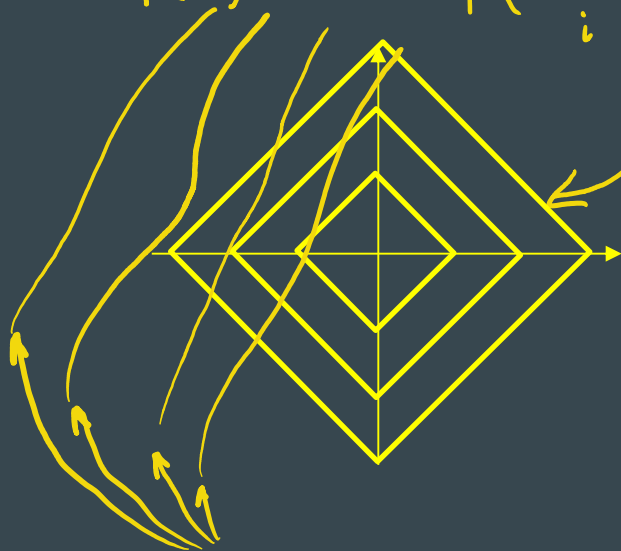


Линии уровня
функции потерь

Регуляризации и их свойства

L1

$$p(w) = \alpha \cdot \exp(-\sum_i |w_i|) \Leftrightarrow + \lambda \sum_i |w_i|$$



Линии уровня регуляризации

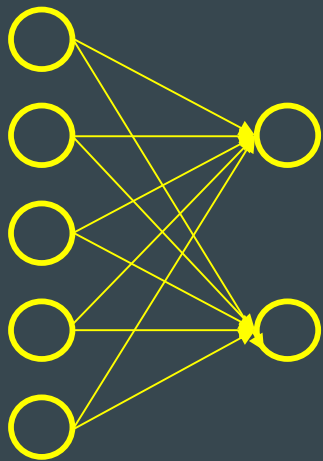
L1 регуляризация неявно отбирает значимые признаки

Линии уровня
функции потерь

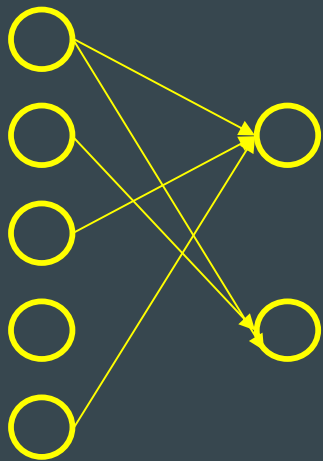
Как можно регуляризовать

- Веса
- Выходы из слоя

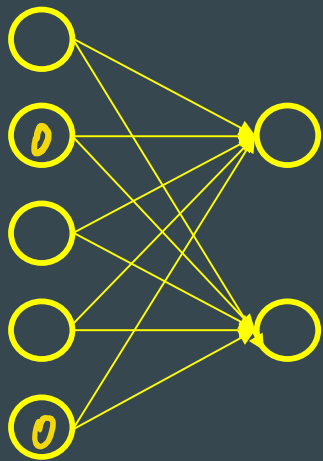
DropOut



DropOut, вариант 1 drop connection



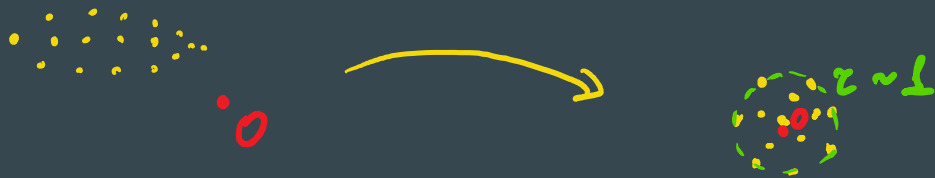
DropOut, вариант 2



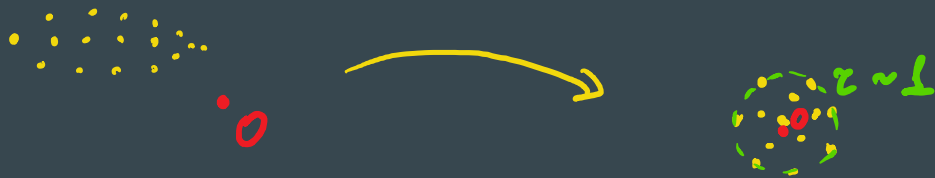
Нормализация данных



Нормализация данных

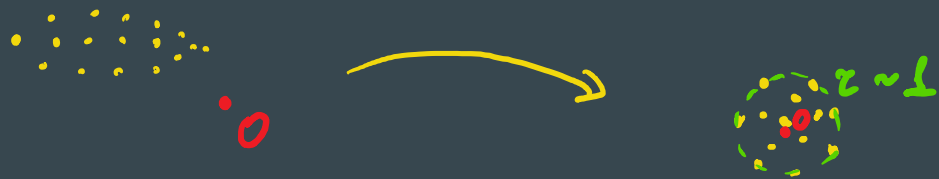


Нормализация данных



$$\vec{x} = \frac{\vec{x} - \vec{\mu}}{\sigma}$$

BatchNorm



$$\bigcirc \rightarrow \vec{y}$$

$$\tilde{y} = \frac{\vec{y} - \mu_y}{\sigma_y}$$

BatchNorm



$$\bigcirc \rightarrow \vec{y}$$

$$\tilde{y} = \frac{\vec{y} - \mu_y}{\sigma_y}$$

Все результаты работы операции
по батчу

BatchNorm

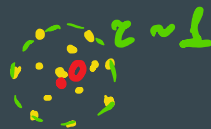


$$\bigcirc \rightarrow \vec{y}$$

$$\tilde{y} = \frac{\vec{y} - \mu_y}{\sigma_y}$$

Среднее и стандартное отклонение по батчу

BatchNorm

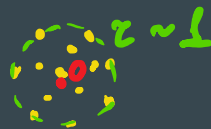
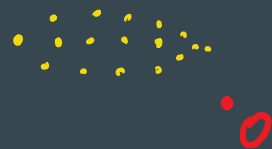


Тренируемые параметры

$$\mathbf{O} \rightarrow \vec{y}$$

$$\tilde{y} = \frac{\vec{y} - \mu_y}{\sigma_y} a + b$$

BatchNorm, inference



Скользящие
средние по батчам

$$\bigcirc \rightarrow \vec{y}$$

$$\tilde{y} = \frac{\vec{y} - \hat{\mu}_y}{\hat{\sigma}_y} a + b$$

Итоги

- Регуляризация
- L2, L1
- Dropout
- BatchNorm