

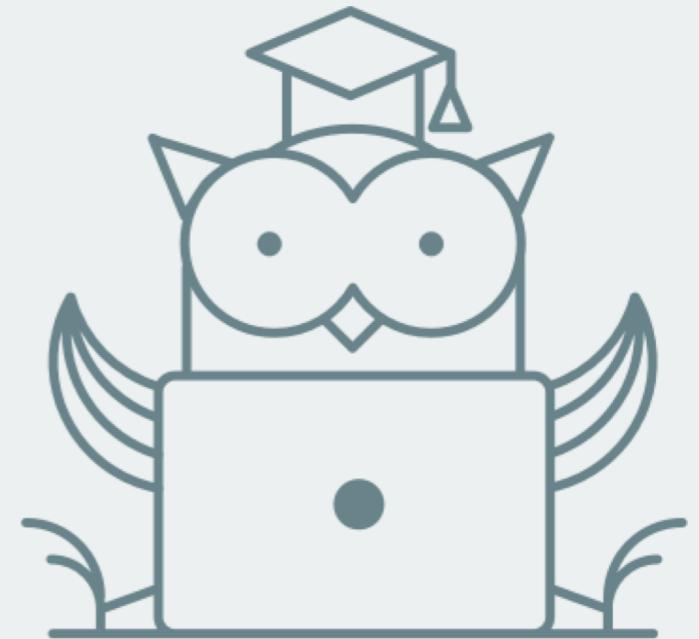
O ·T· U S

ОНЛАЙН-ОБРАЗОВАНИЕ

Переобучение и регуляризация

Как и чему учить нейронные сети

Артур Кадурин
Преподаватель



План на сегодня

1. Валидация
2. Переобучение и регуляризация
3. Нормализация мини-батчами и дропаут
4. Практика



Dataset

Training

Testing

Holdout Method

Вопросы:

1. Зачем нужно тестовое множество?



Dataset

Training

Testing

Holdout Method

Вопросы:

- 1. Зачем нужно тестовое множество?**
- 2. Как выбрать тестовое множество?**



Dataset

Training

Testing

Holdout Method

Вопросы:

- 1. Зачем нужно тестовое множество?**
- 2. Как выбрать тестовое множество?**
- 3. Как определить размер тестового множества?**



Dataset

Training

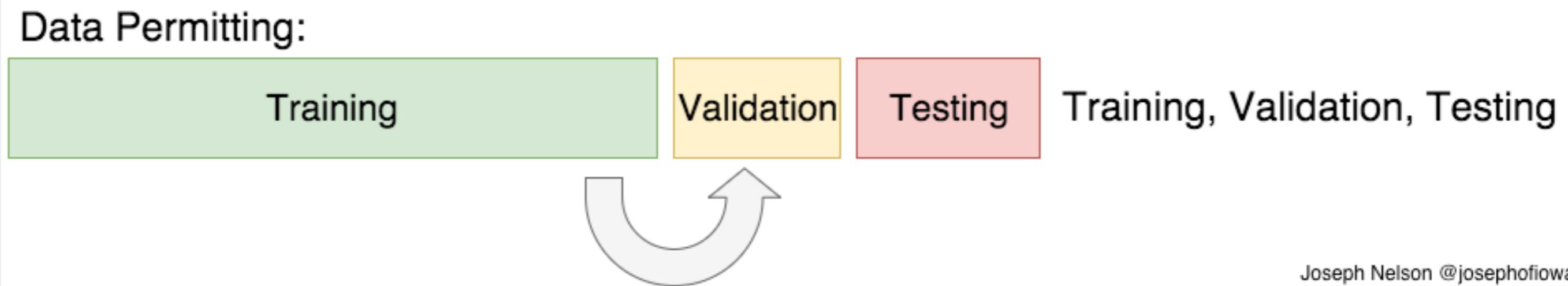
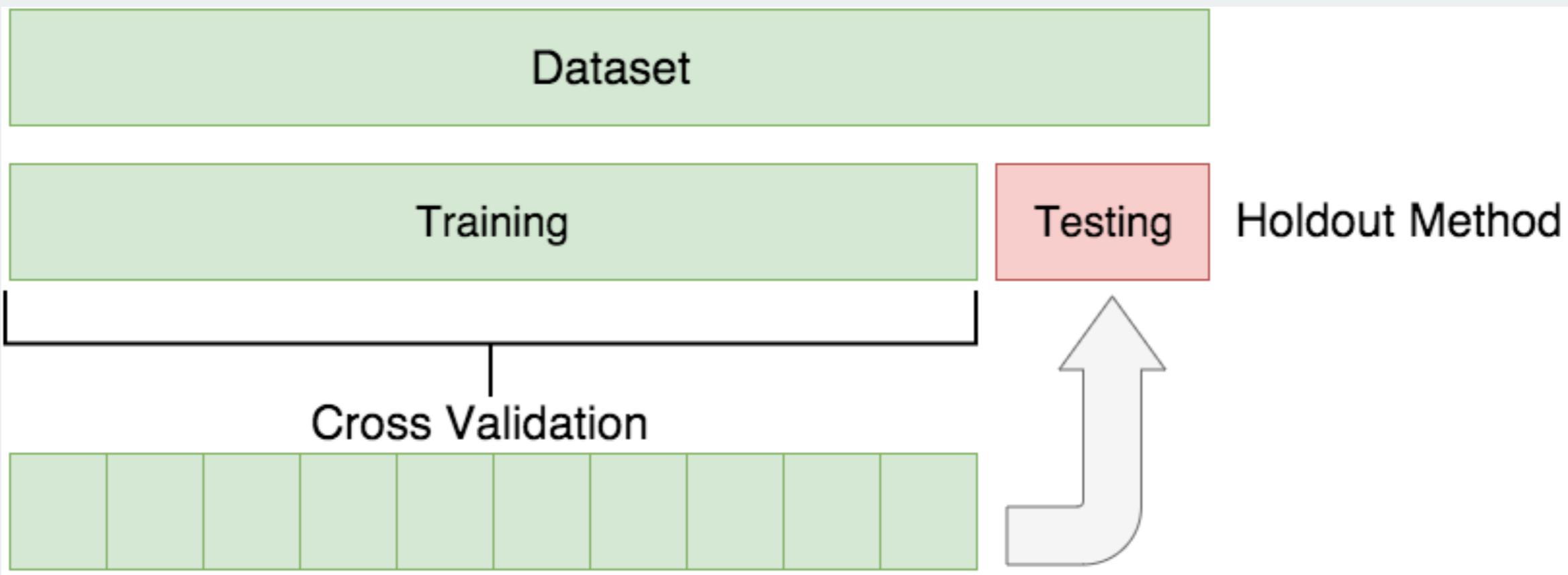
Testing

Holdout Method

Вопросы:

- 1. Зачем нужно тестовое множество?**
- 2. Как выбрать тестовое множество?**
- 3. Как определить размер тестового множества?**
- 4. Как сравнивать два разных класса моделей?**





Joseph Nelson @josephofiowa



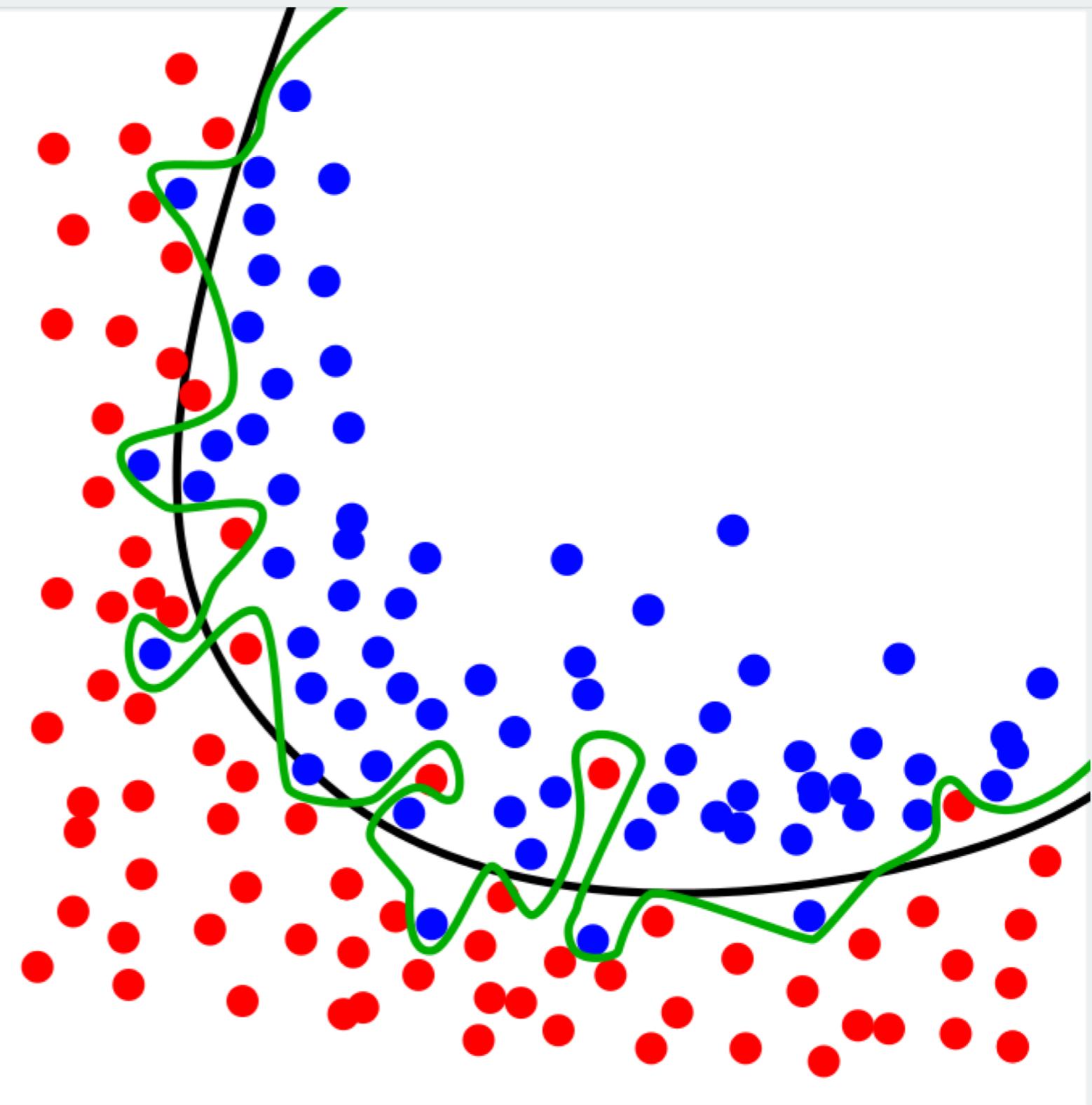
План на сегодня

1. Валидация
- 2. Переобучение и регуляризация**
3. Нормализация мини-батчами и дропаут
4. Практика



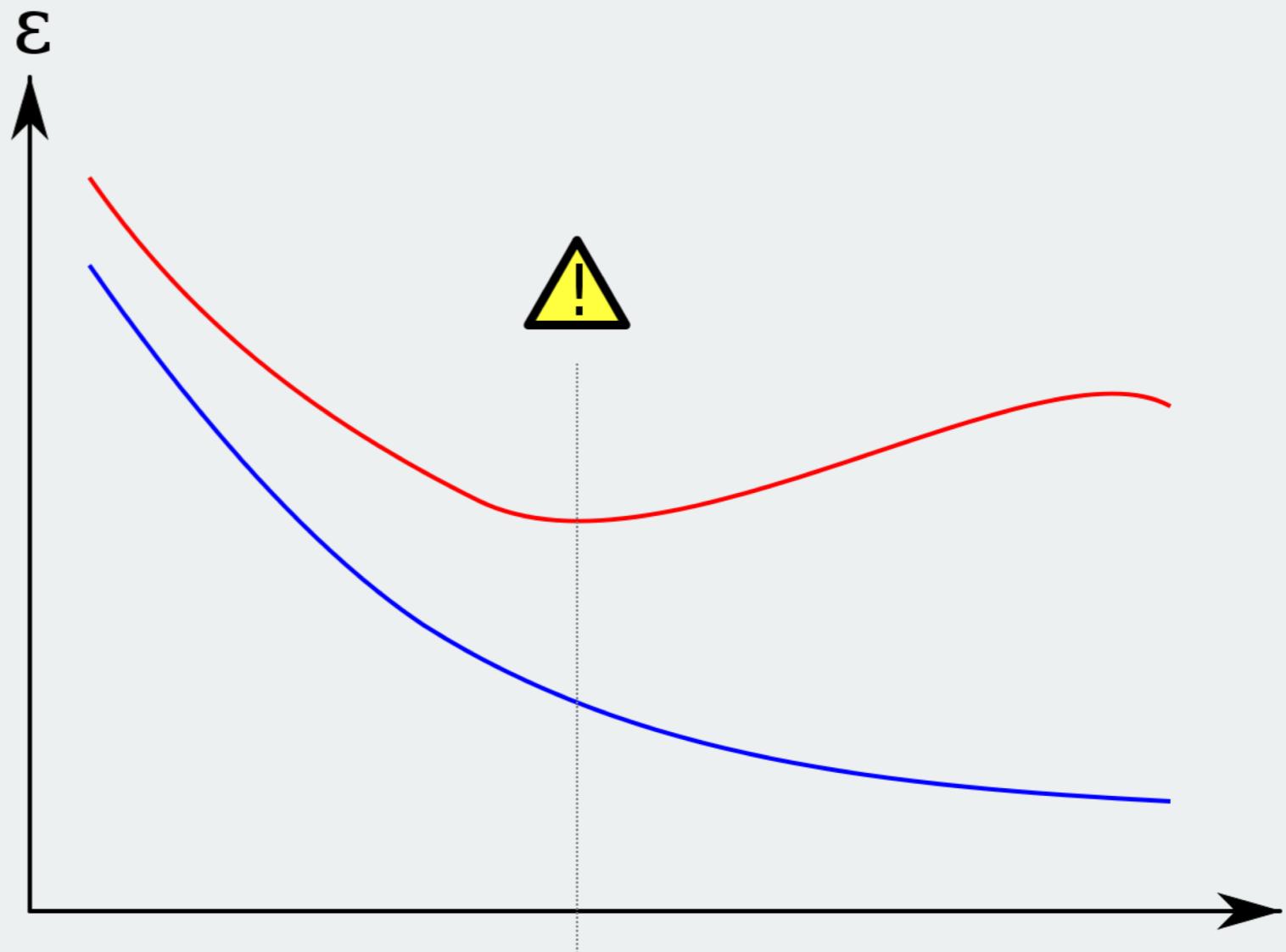
Переобучение.

1. Смещенная выборка
2. Недостаточная выборка



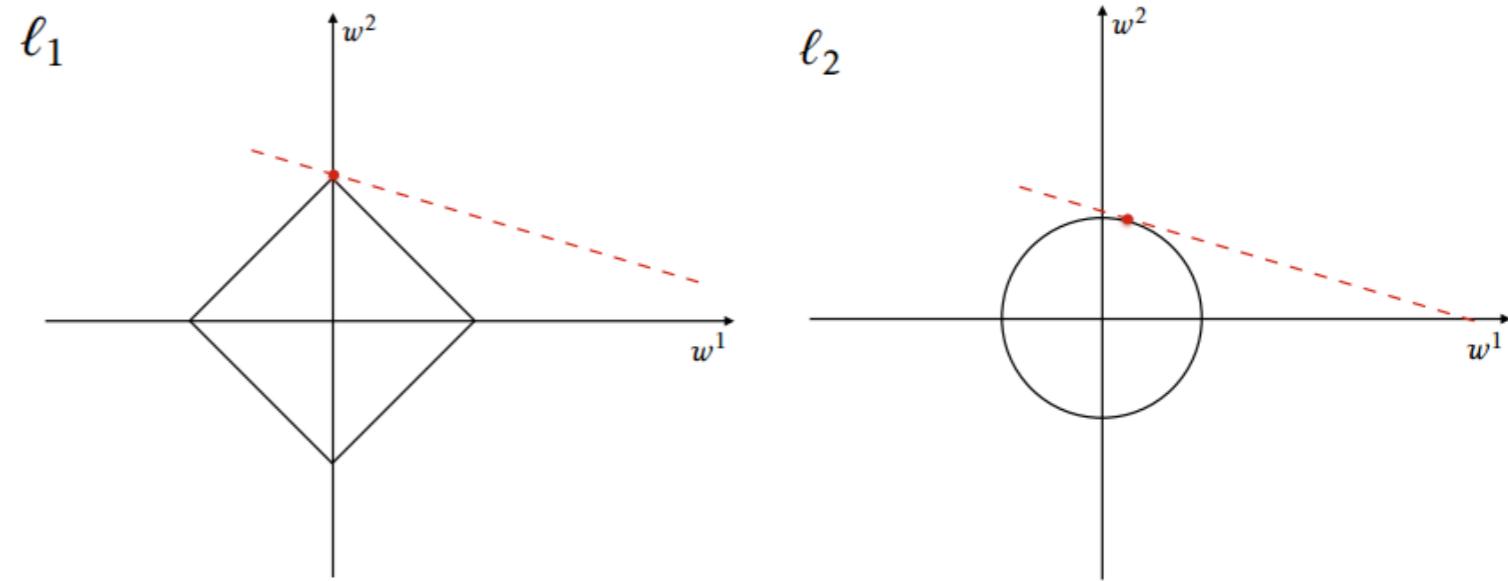
Способы борьбы:

1. Ранняя остановка



Способы борьбы:

1. Ранняя остановка
2. Регуляризация:



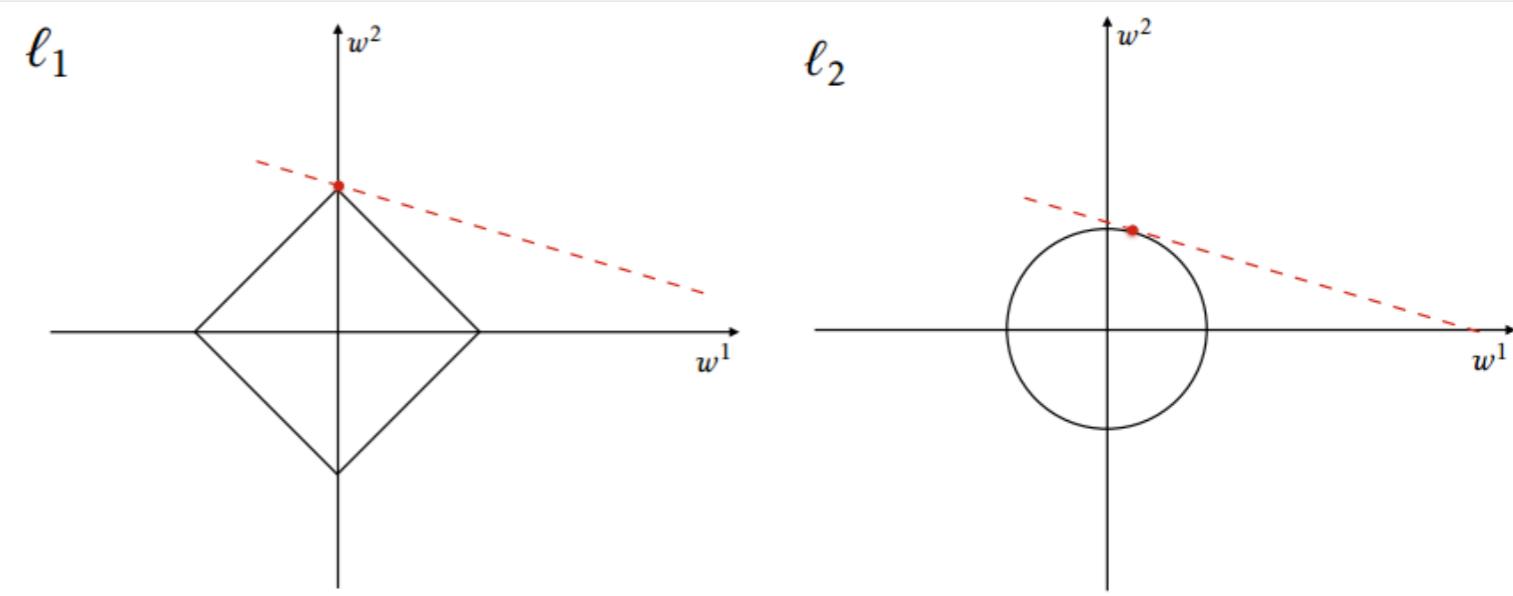
Регуляризация — это наложение дополнительных ограничений.

Как мы можем это сделать?



Способы борьбы:

1. Ранняя остановка
2. Регуляризация:



Регуляризация — это наложение дополнительных ограничений.

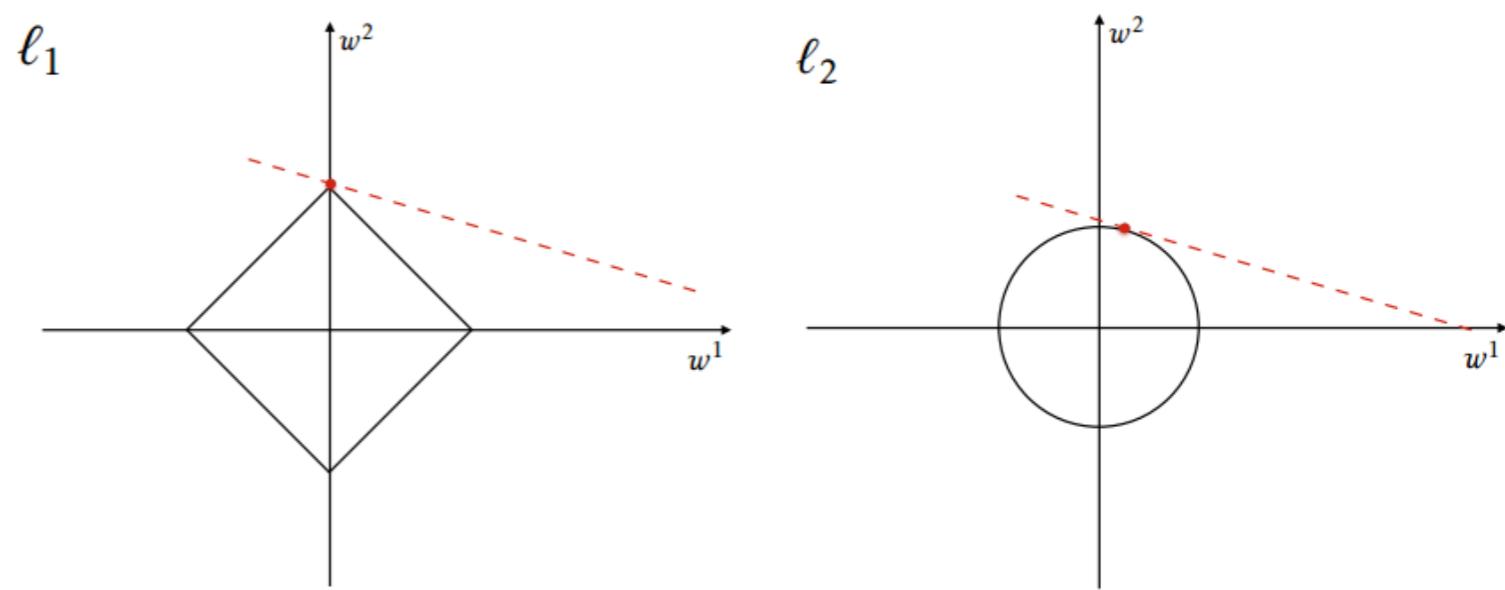
Как мы можем это сделать?

$$\min_f \mathcal{L}(f(x), y) \Rightarrow \min_f \left[\sum_i \mathcal{L}(f(x_i), y_i) + \lambda R(f) \right]$$



Способы борьбы:

1. Ранняя остановка
2. Регуляризация:



Регуляризация – это наложение дополнительных ограничений.

Как мы можем это сделать?

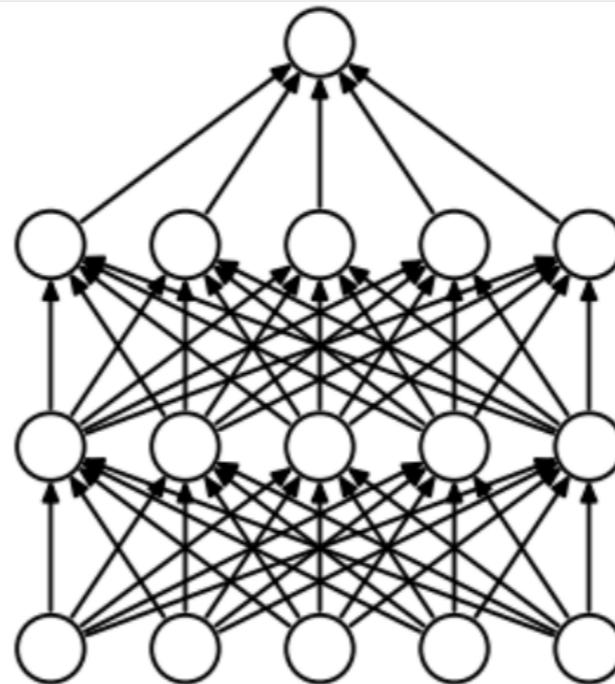
$$\min_f \mathcal{L}(f(x), y) \Rightarrow \min_f \left[\sum_i \mathcal{L}(f(x_i), y_i) + \lambda R(f) \right]$$

$$R(f(x|w)) = \|w\|_1 \qquad \qquad R(f(x|w)) = \|w\|_2^2$$

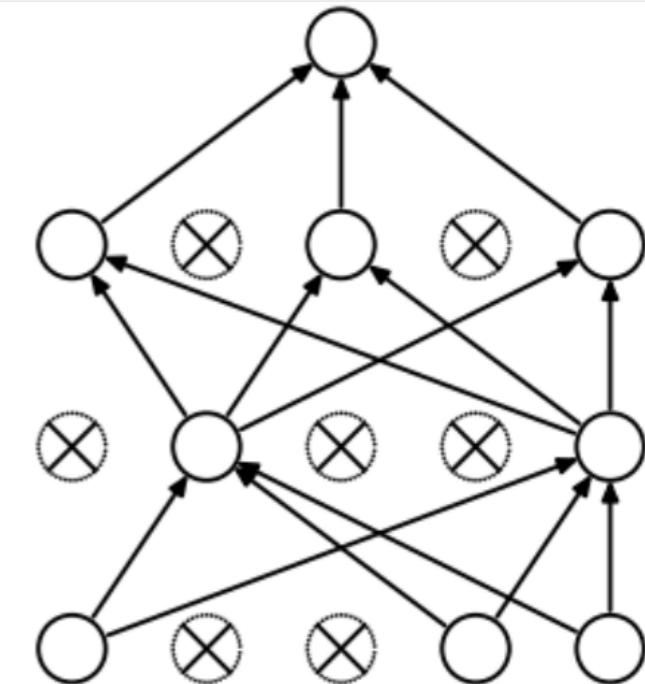


Способы борьбы:

1. Ранняя остановка
2. Регуляризация
3. Специальные слои



(a) Standard Neural Net



(b) After applying dropout.

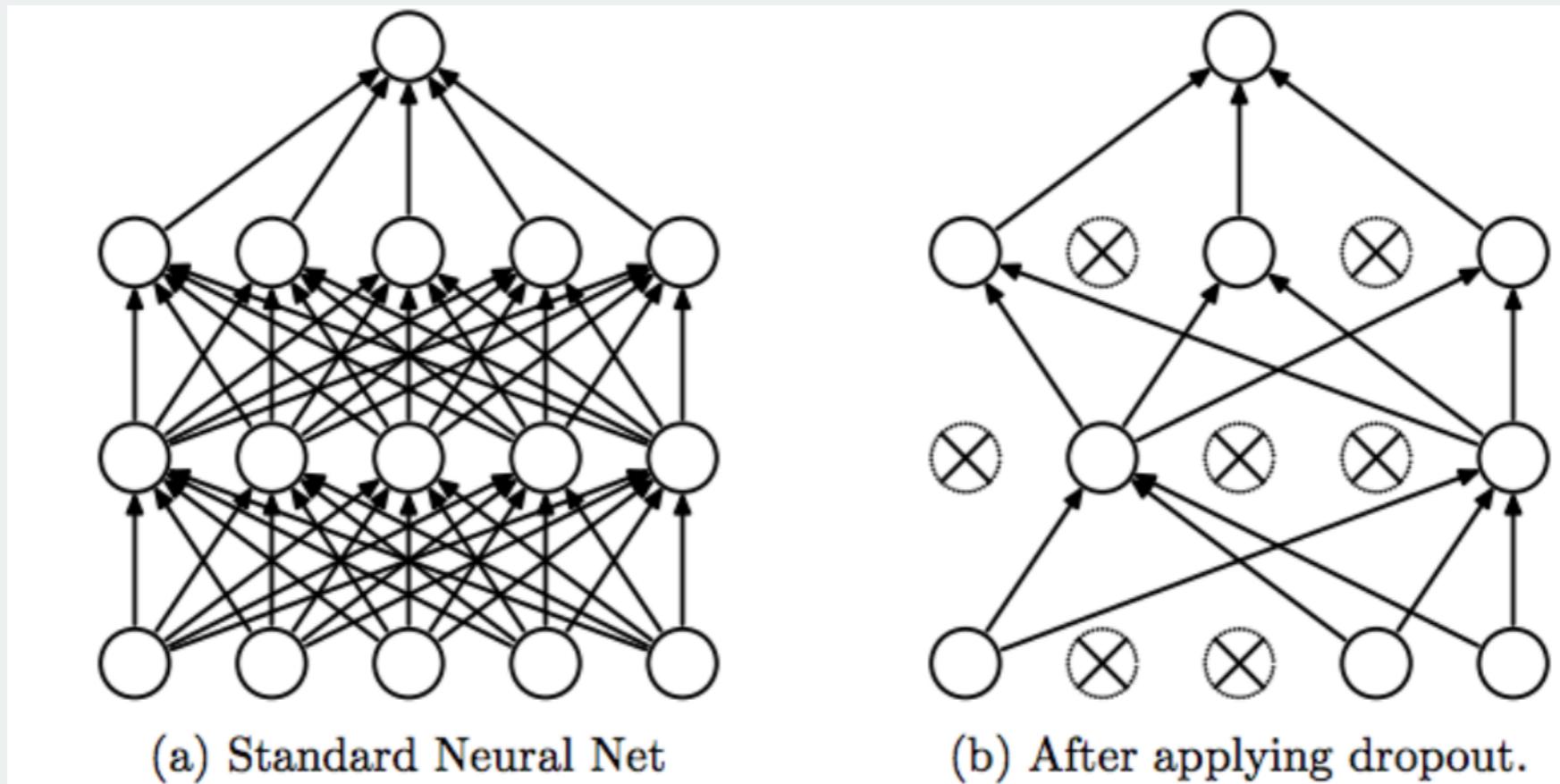


План на сегодня

1. Валидация
2. Переобучение и регуляризация
- 3. Нормализация мини-батчами и дропаут**
4. Практика



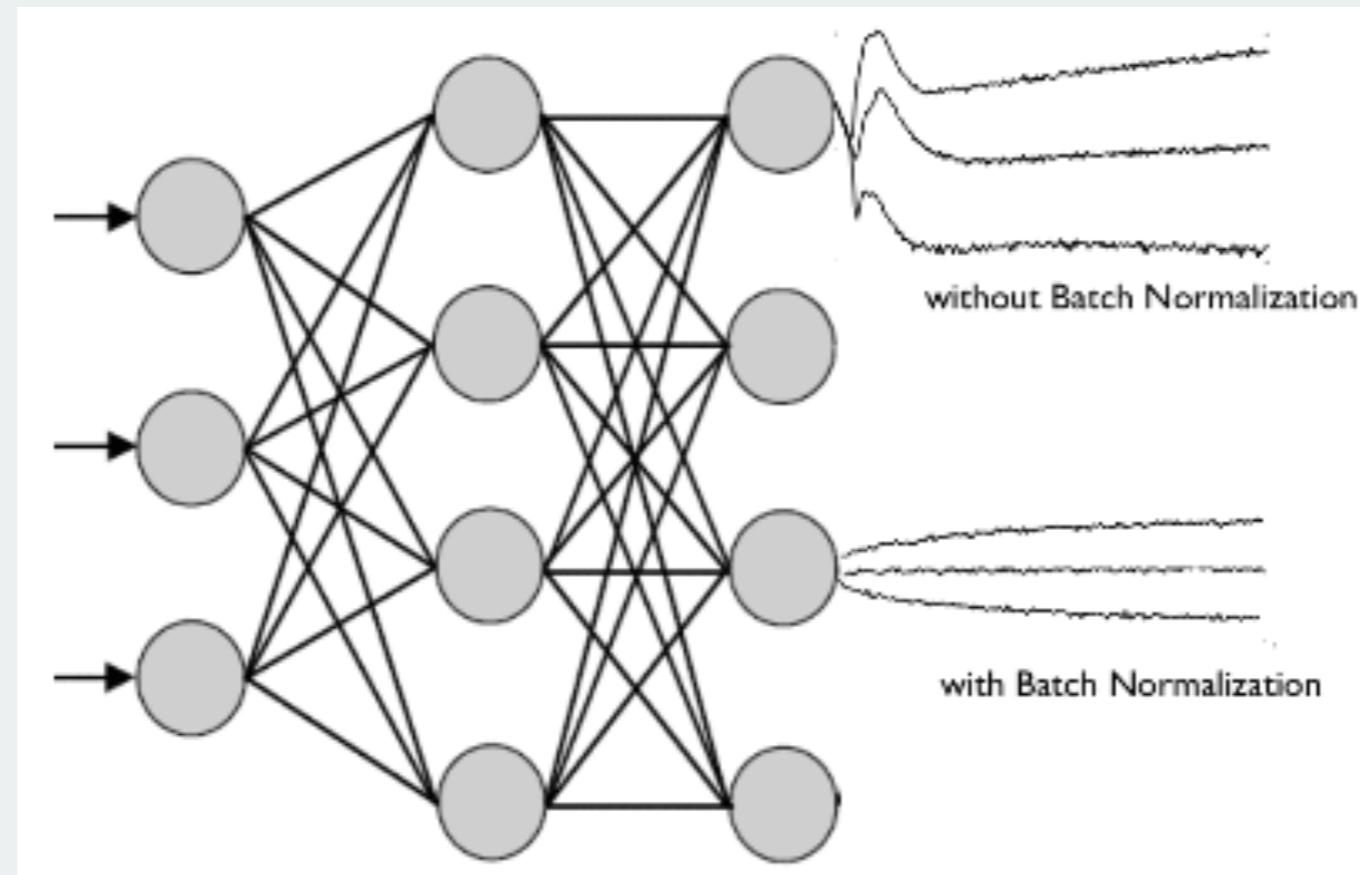
Дропаут



В процессе обучения нейронной сети мы можем, как бы, временно «выключать» часть нейронов.



Батч-нормализация



В процессе обучения нейронной сети мы можем нормировать активации каждого слоя используя статистики текущего батча.



Батч-нормализация

$$\boldsymbol{x} = (x_0, x_1, \dots, x_n), \hat{x}_k = \frac{x_k - \mathbb{E}[x_k]}{\sqrt{Var(x_k)}}$$



Батч-нормализация

$$\boldsymbol{x} = (x_0, x_1, \dots, x_n), \hat{x}_k = \frac{x_k - \mathbb{E}[x_k]}{\sqrt{Var(x_k)}}$$

$$y_k = \gamma_k \hat{x}_k + \beta_k = \gamma_k \frac{x_k - \mathbb{E}[x_k]}{\sqrt{Var(x_k)}} + \beta_k$$



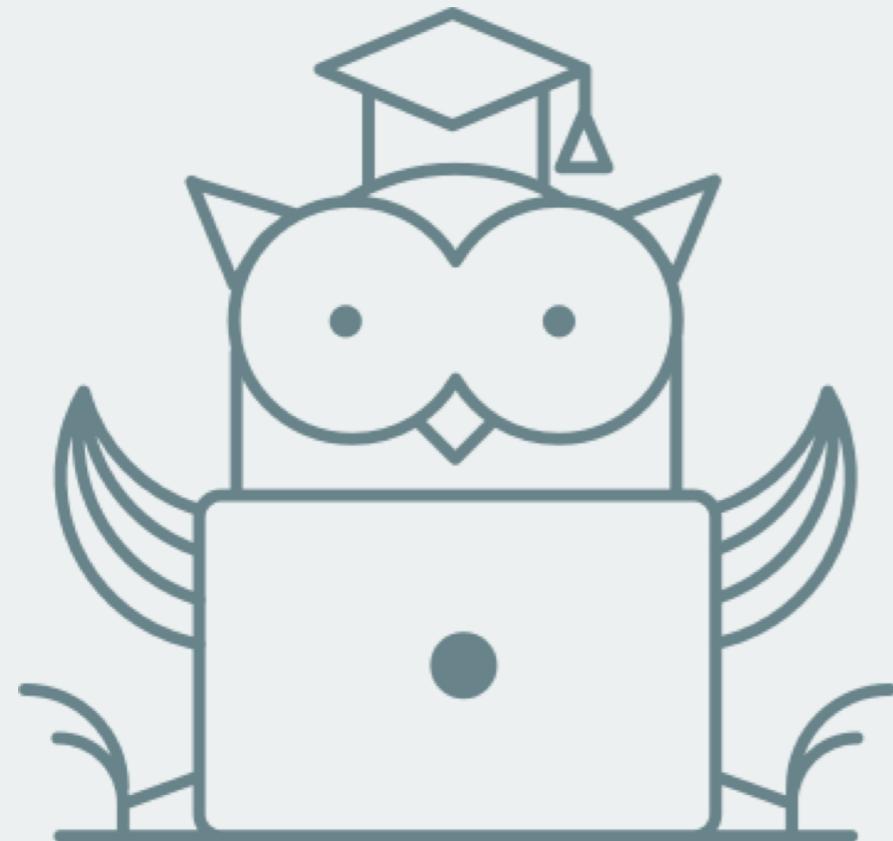
Батч-нормализация

$$\mathbf{x} = (x_0, x_1, \dots, x_n), \hat{x}_k = \frac{x_k - \mathbb{E}[x_k]}{\sqrt{Var(x_k)}}$$

$$y_k = \gamma_k \hat{x}_k + \beta_k = \gamma_k \frac{x_k - \mathbb{E}[x_k]}{\sqrt{Var(x_k)}} + \beta_k$$

If $\gamma_k = \sqrt{Var(x_k)}$ and $\beta_k = \mathbb{E}[x_k]$, then $y_k = x_k$





Спасибо
за внимание!