



ОНЛАЙН-ОБРАЗОВАНИЕ

Переобучение и регуляризация

Как и чему учить нейронные сети

Артур Кадури
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

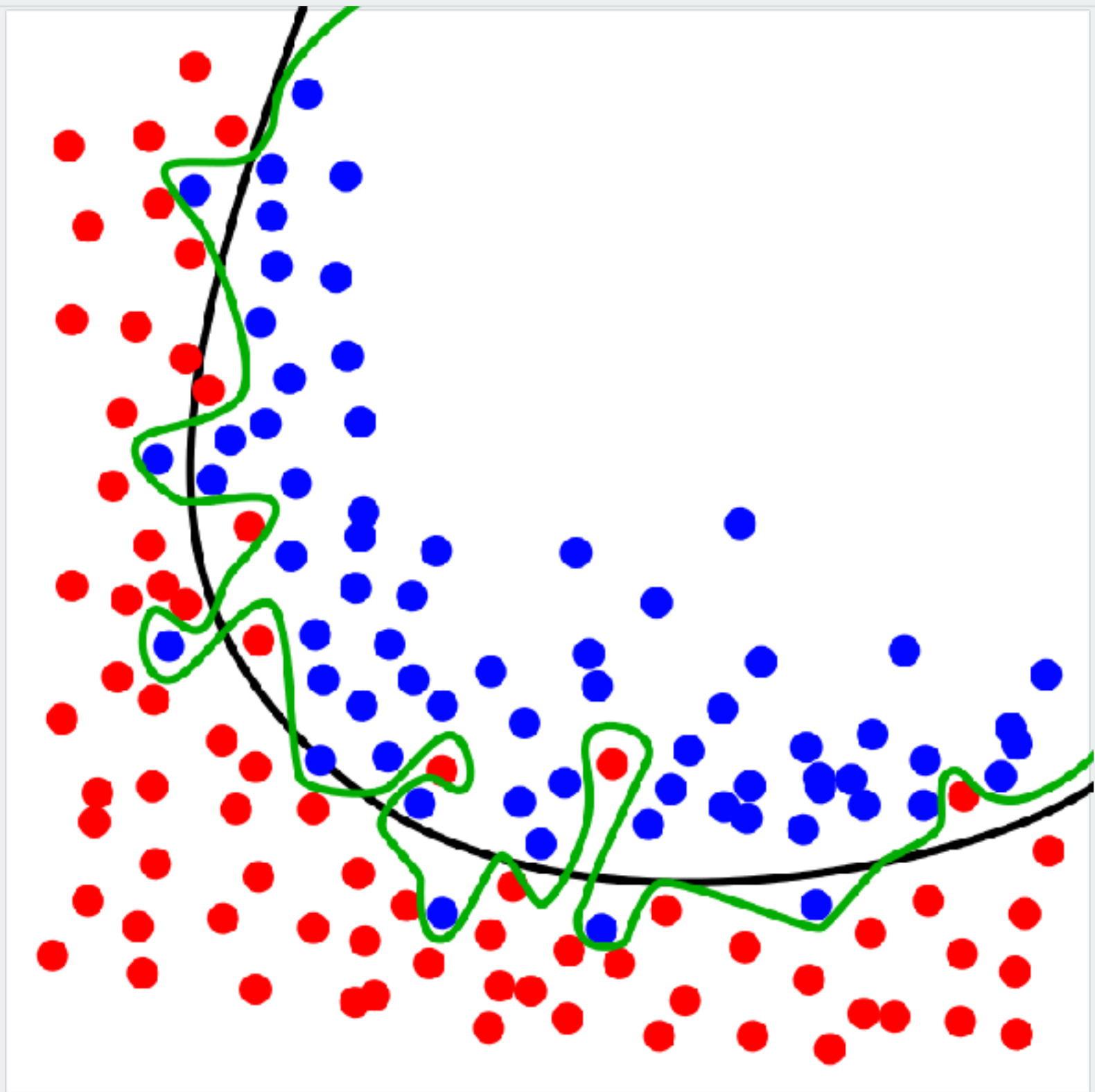
Вопросы:

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



Переобучение.

1. Смещенная выборка
2. Недостаточная выборка



Dataset

Training

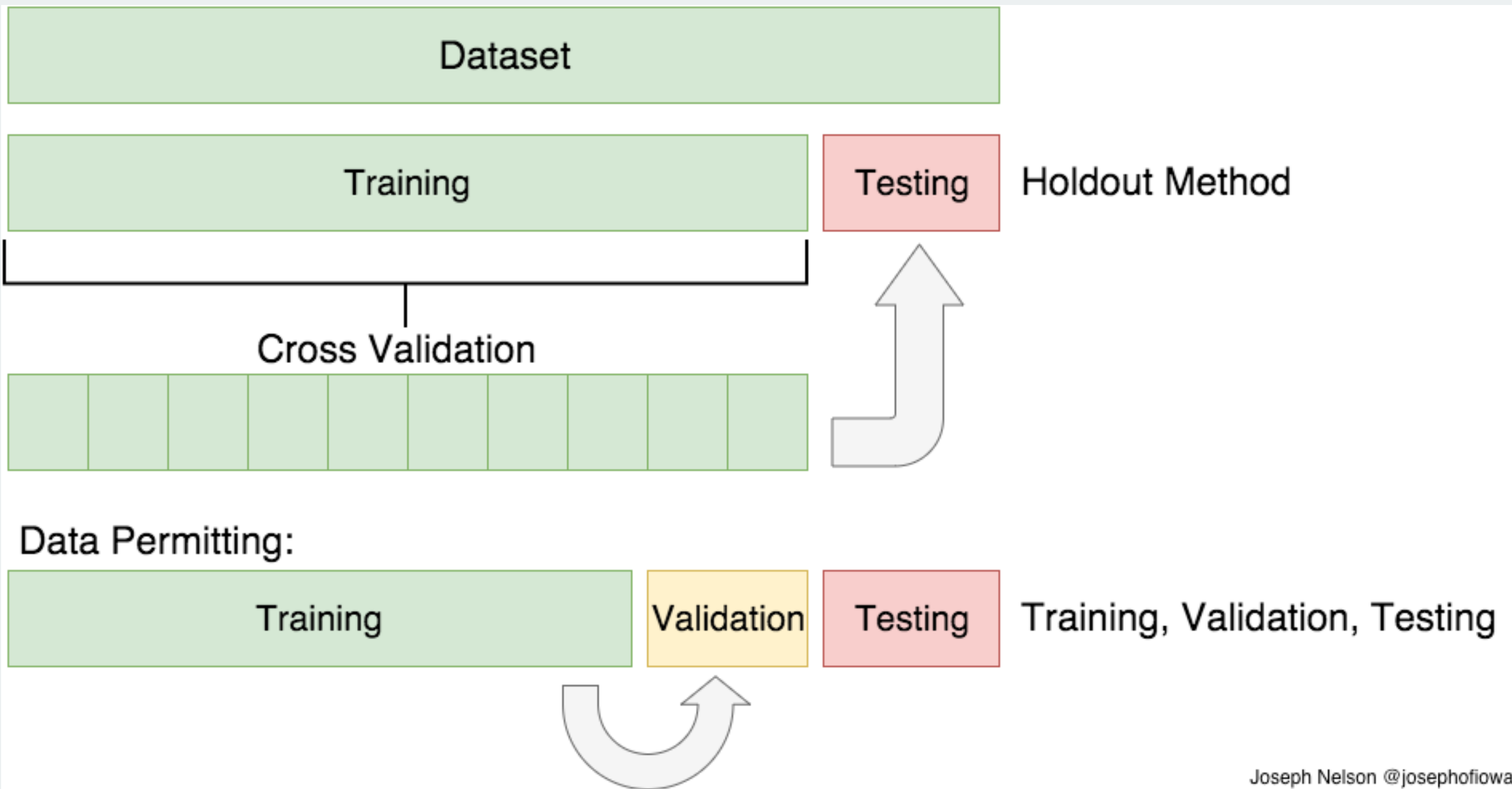
Testing

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?





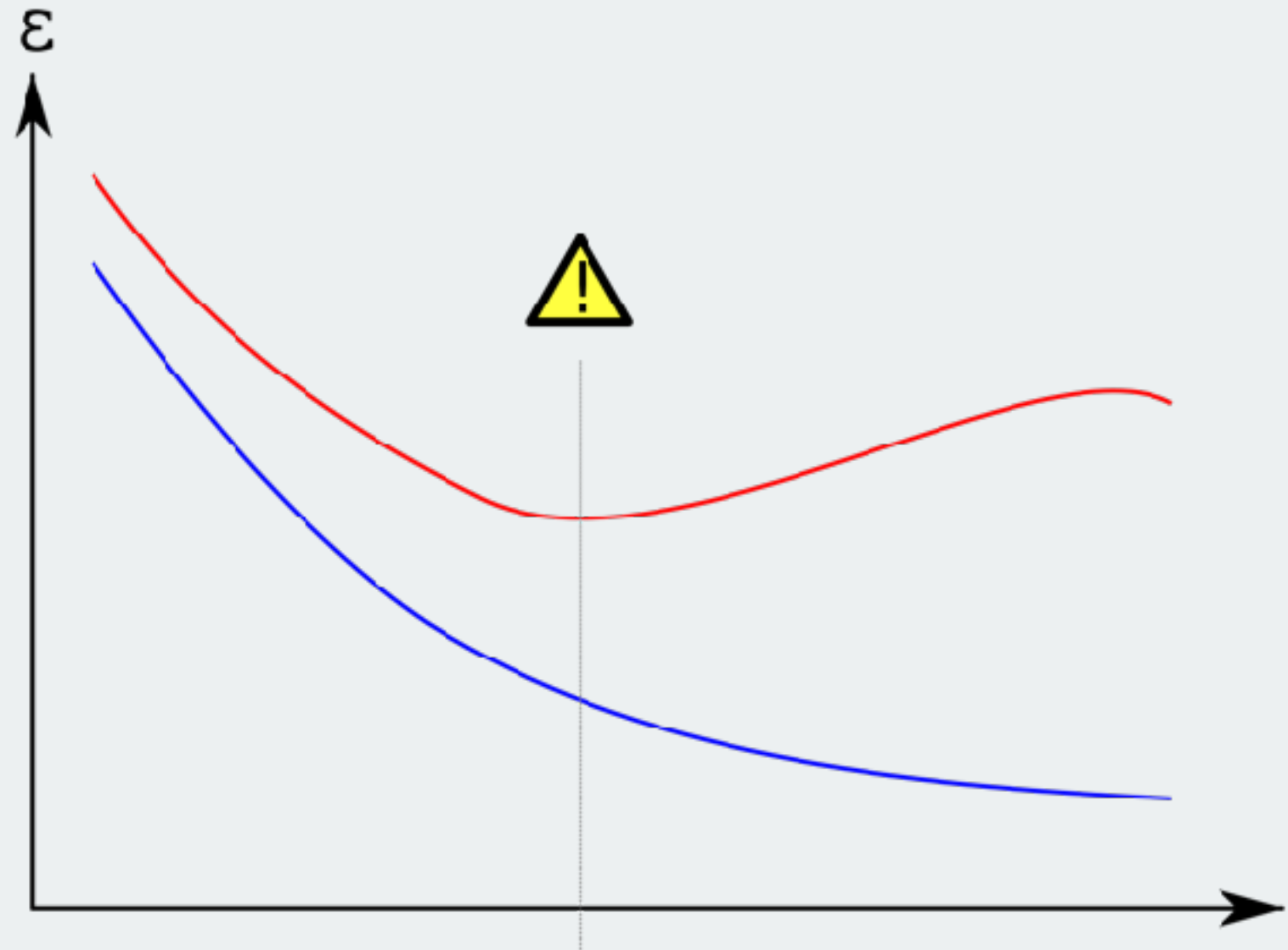
План на сегодня

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



Способы борьбы:

1. Ранняя остановка



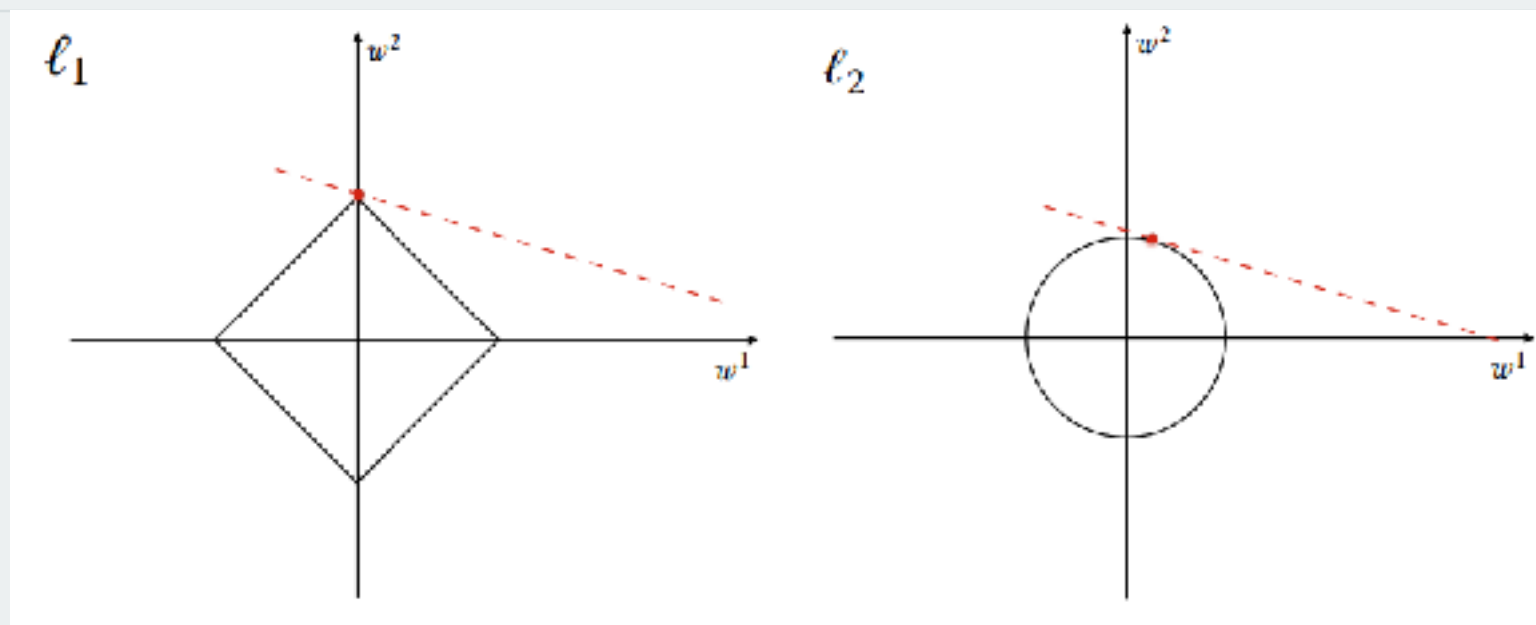
Способы борьбы:

1. Ранняя остановка

2. Регуляризация:

Регуляризация — это наложение дополнительных ограничений.

Как мы можем это сделать?



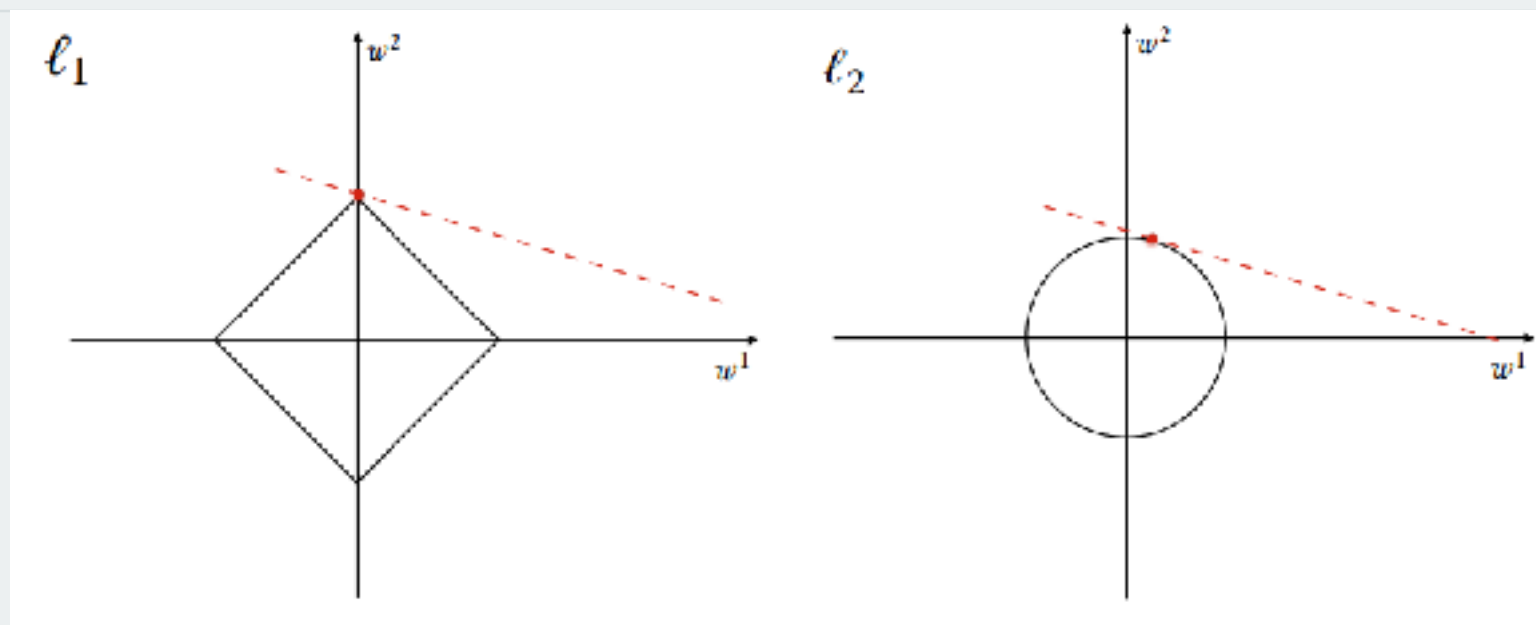
Способы борьбы:

1. Ранняя остановка

2. Регуляризация:

Регуляризация — это наложение дополнительных ограничений.

Как мы можем это сделать?



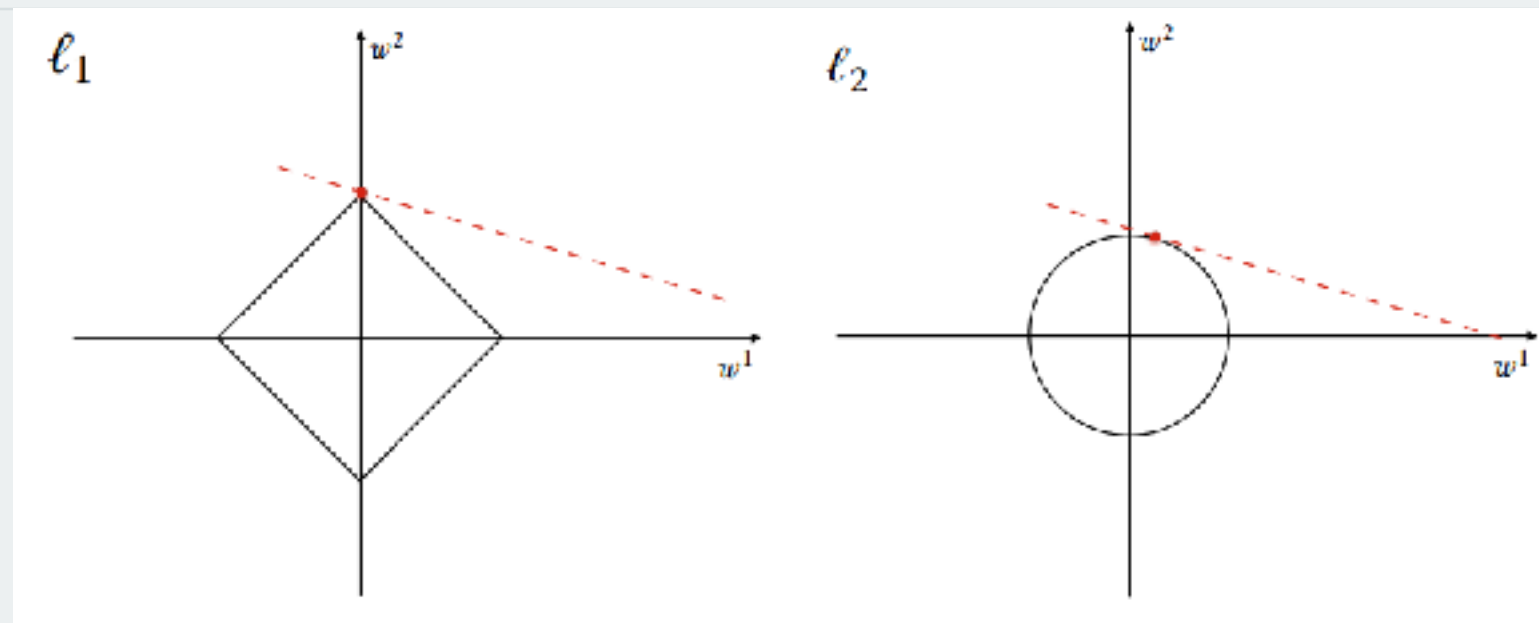
$$\min_f \mathcal{L}(f(\mathbf{x}), \mathbf{y}) \Rightarrow \min_f \left[\sum_i \mathcal{L}(f(x_i), y_i) + \lambda R(f) \right]$$



Способы борьбы:

1. Ранняя остановка

2. Регуляризация:



Регуляризация — это наложение дополнительных ограничений.

Как мы можем это сделать?

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

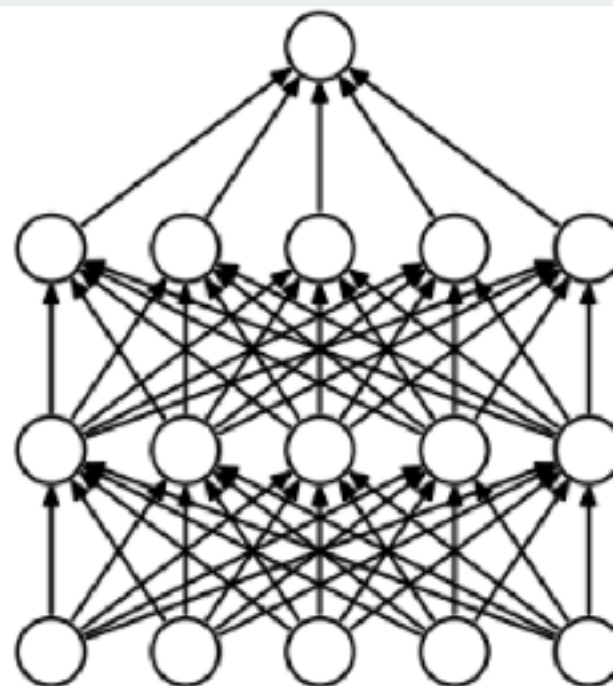
$$R(f(\mathbf{x}|\mathbf{w})) = \|\mathbf{w}\|_1$$

$$R(f(\mathbf{x}|\mathbf{w})) = \|\mathbf{w}\|_2^2$$

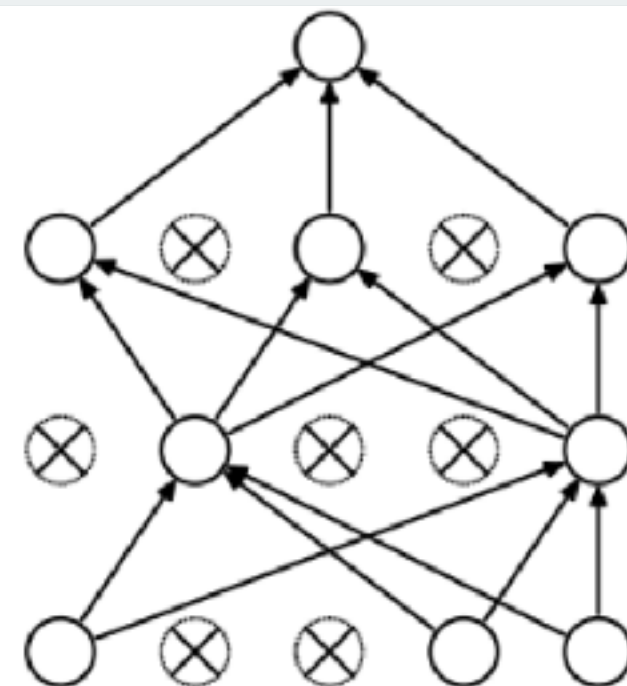


Способы борьбы:

1. Ранняя остановка
2. Регуляризация
3. Специальные слои



(a) Standard Neural Net



(b) After applying dropout.

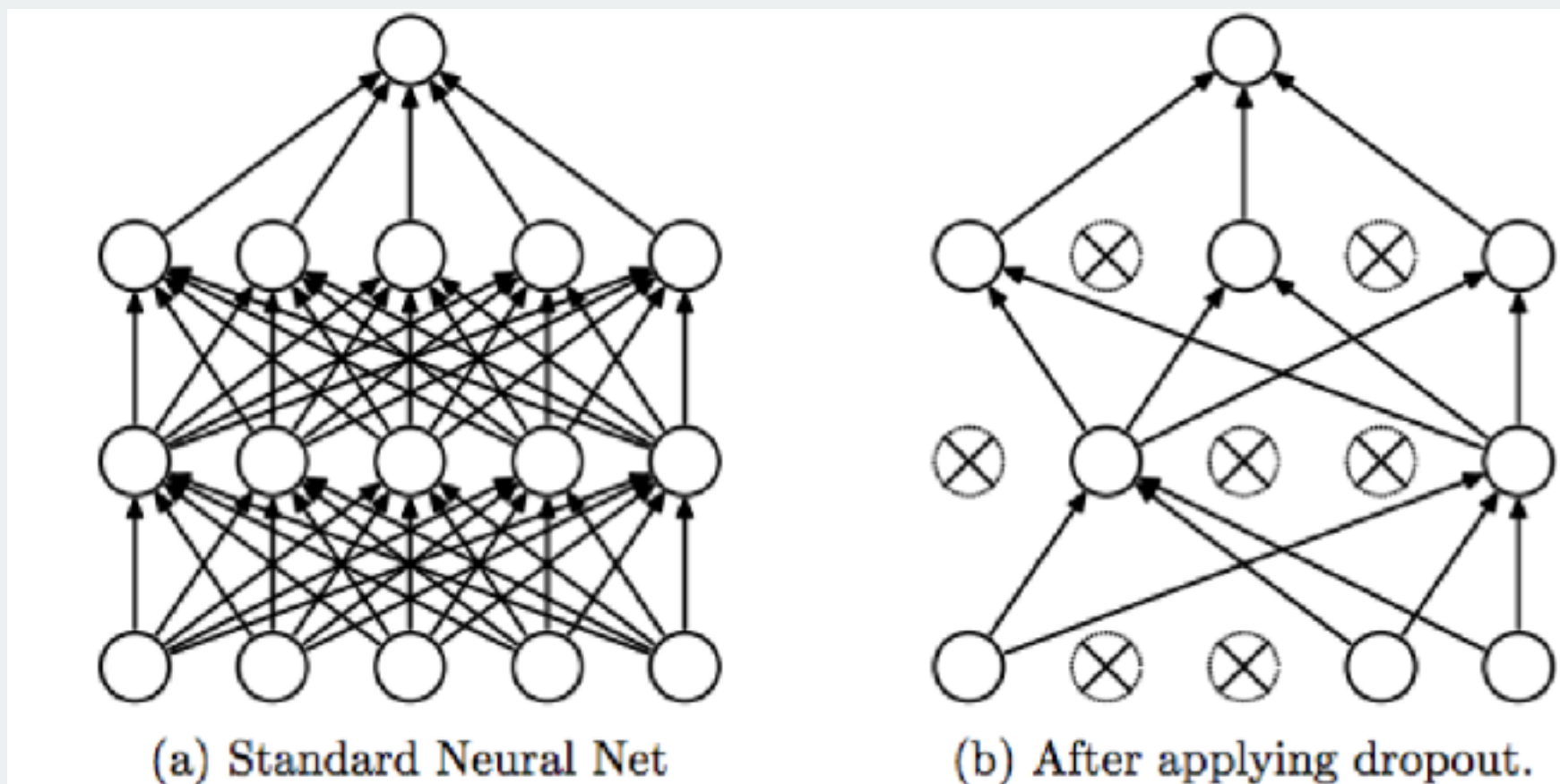


План на сегодня

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



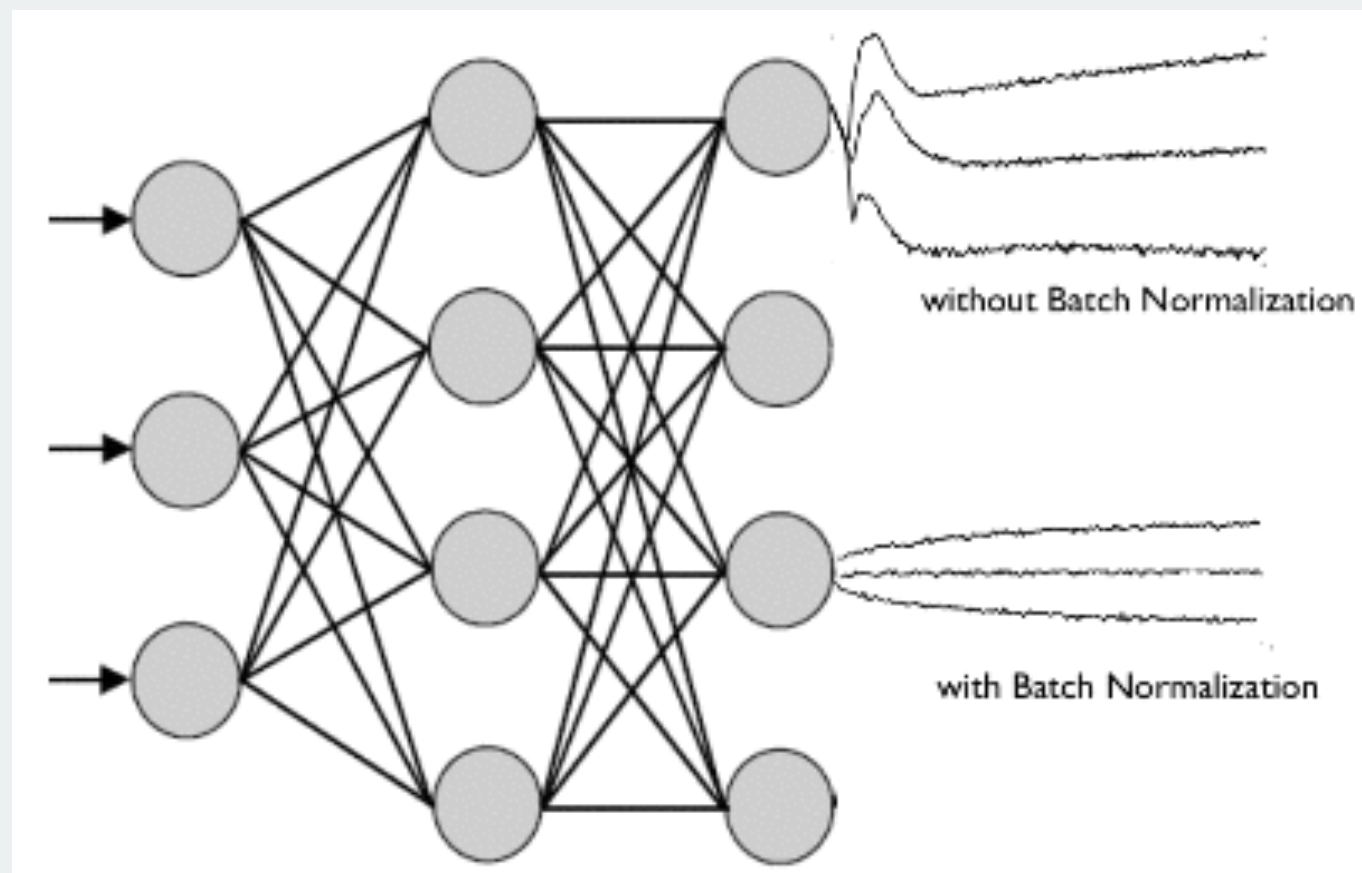
Дропаут



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

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



Батч-нормализация

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

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

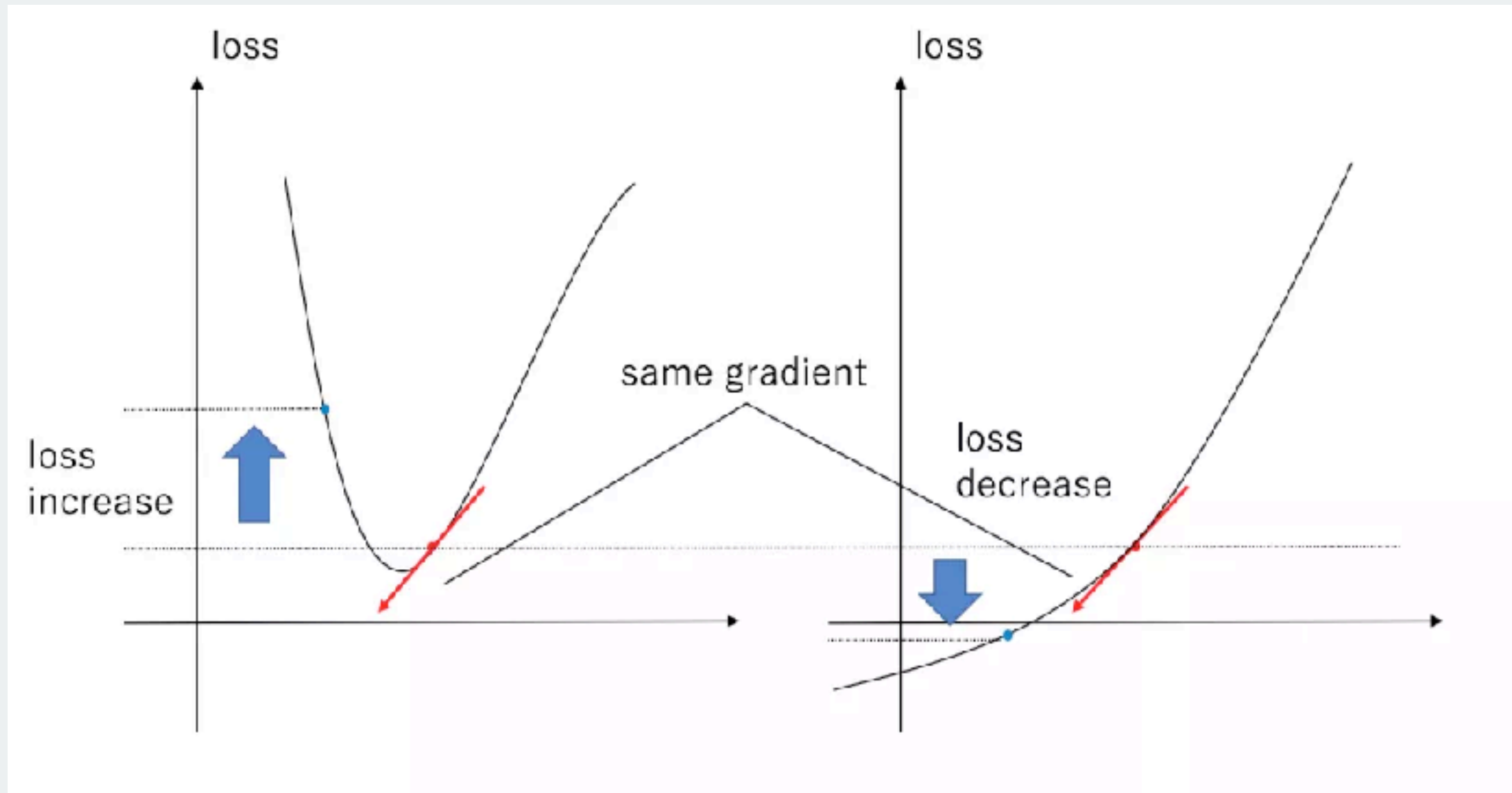


Батч-нормализация

- (1) Even if the mean and variance are constant, the distribution of activations can still change. Why are the mean and variance so important?
- (2) If we introduce γ and β , the mean and variance will deviate from 0 and 1 anyway. What then, is the point of batch normalization?



Батч-нормализация



- <https://mlexplained.com/2018/01/10/an-intuitive-explanation-of-why-batch-normalization-really-works-normalization-in-deep-learning-part-1/>
- <https://www.youtube.com/watch?v=Xogn6veSyxA&feature=youtu.be&t=325>





Спасибо
за внимание!