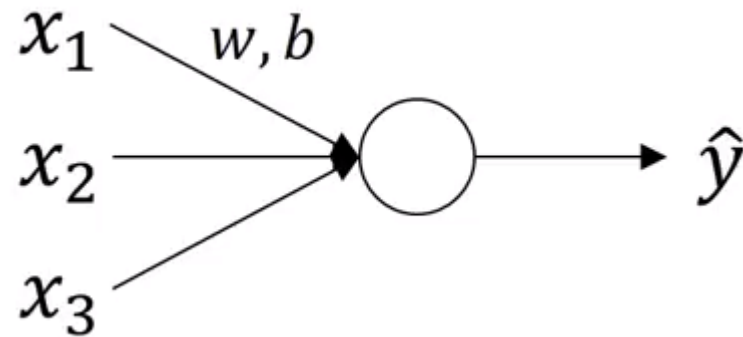


Batch Normalization

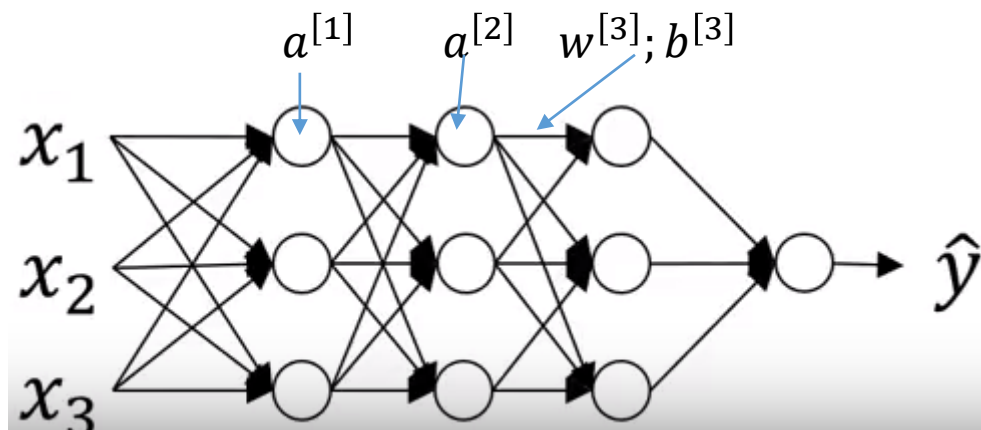
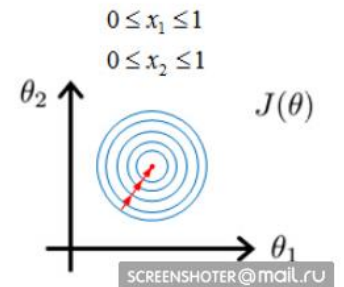
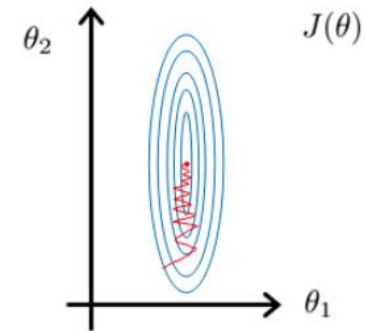
Нормализация активаций



$$\mu_B \leftarrow \frac{1}{m} \sum_{i=1}^m x_i$$

$$\sigma_B^2 \leftarrow \frac{1}{m} \sum_{i=1}^m (x_i - \mu_B)^2$$

$$\hat{x}_i \leftarrow \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}}$$



Нормализуем $a^{[2]}$, что бы $w^{[3]}; b^{[3]}$ тренировались быстрее нормализуем вход функции активации $z^{[2]}$

$$z^{[2]} = \sum_i a^{[1]} \dots w^{[2]} \dots$$

Нормализуем $a^{[l]}$

Для второго слоя для каждого промежуточного значения z для всего батча: z^1, \dots, z^m

- $\mu = \frac{1}{m} \sum z^i$
 - $\sigma^2 = \frac{1}{m} \sum (z_i - \mu)^2$
 - $z_{norm}^i = \frac{z^i - \mu}{\sqrt{\sigma^2 + \epsilon}}$
 - $\tilde{z}^i = \gamma z_{norm}^i + \beta$
- Обучаемые
параметры

Если:

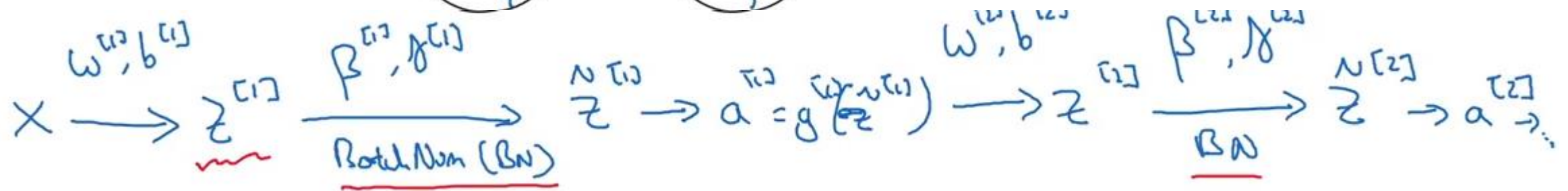
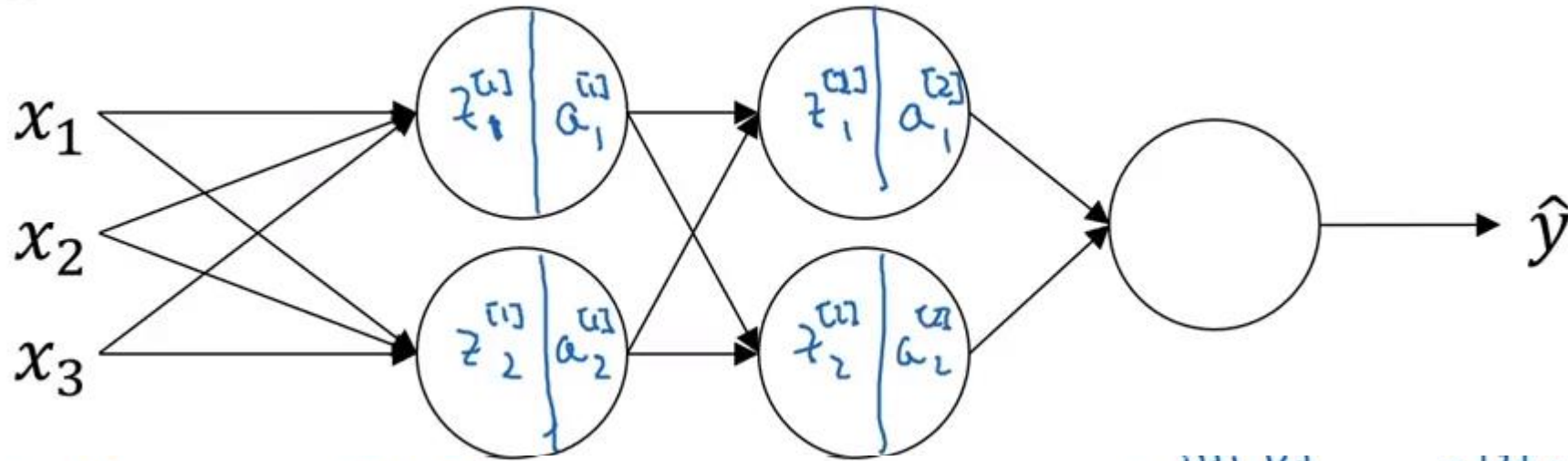
$$\gamma = \sqrt{\sigma^2 + \epsilon}$$

$$\beta = \mu$$

Получаем идентичный мапинг

$$\tilde{z}^i = z^i$$

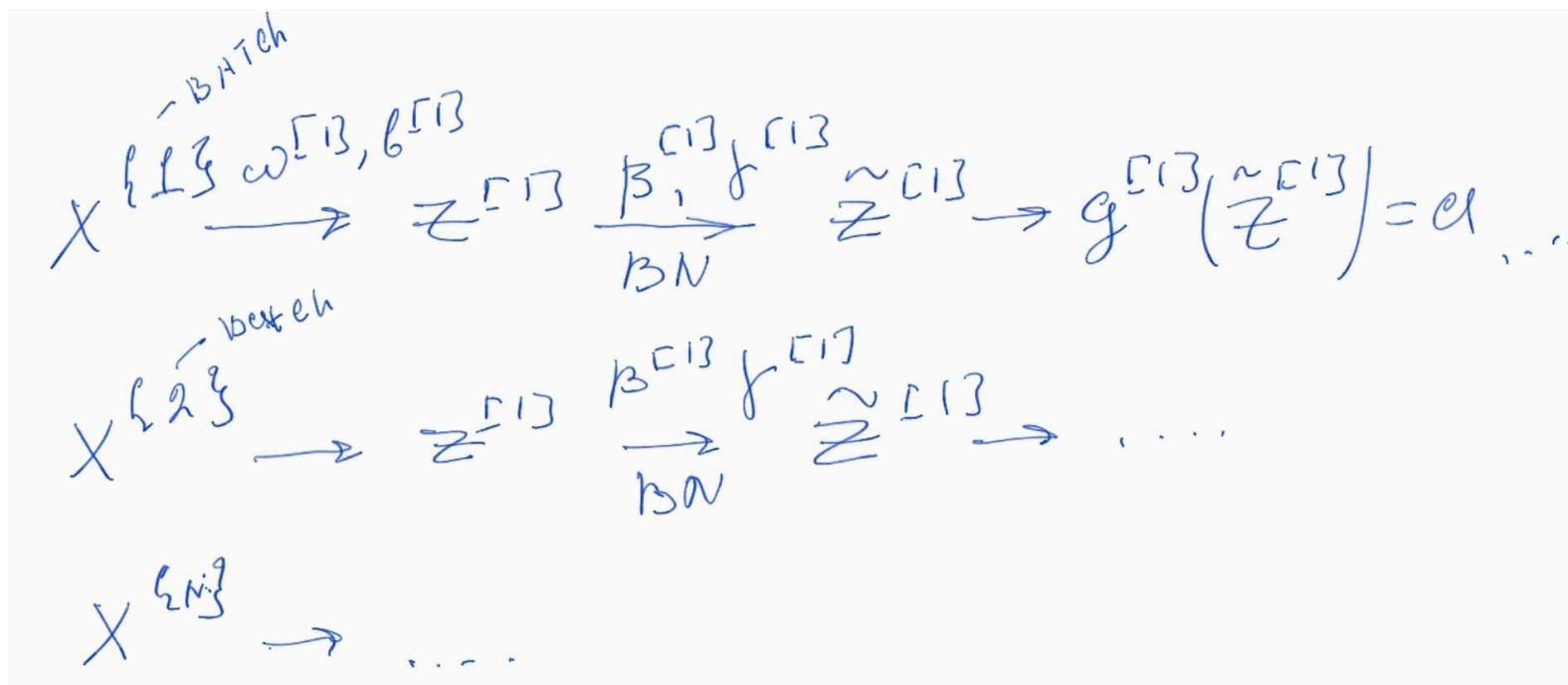
Обучение BN в сети



$$\left. \begin{aligned} &W^{(1)}, \underline{b}^{(1)}, W^{(2)}, \underline{b}^{(2)}, \dots, W^{(L)}, \underline{b}^{(L)} \\ &\rightarrow \underline{\beta}^{(1)}, \underline{\gamma}^{(1)}, \underline{\beta}^{(2)}, \underline{\gamma}^{(2)}, \dots, \underline{\beta}^{(L)}, \underline{\gamma}^{(L)} \end{aligned} \right\} \quad \begin{aligned} &d\beta^{(L)} \\ &\beta = \beta - \alpha d\beta^{(L)} \end{aligned}$$

Параметры

BN для мини-батчей



Обновление весов

$$\left. \begin{aligned} w^{[L]} &:= w^{[L]} - \alpha dw^{[L]} \\ \beta^{[L]} &:= \beta^{[L]} - \alpha d\beta^{[L]} \\ \gamma^{[L]} &:= \dots \end{aligned} \right\}$$

Параметры

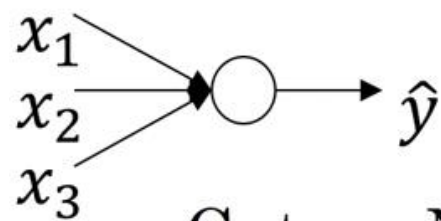
$$w^{[L]}, \cancel{b^{[L]}}, \beta^{[L]}, \gamma^{[L]}$$

$$\rightarrow \underline{z^{[L]}} = w^{[L]} a^{[L-1]} + \cancel{b^{[L]}}$$

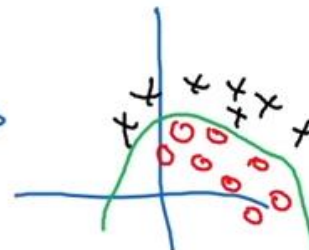
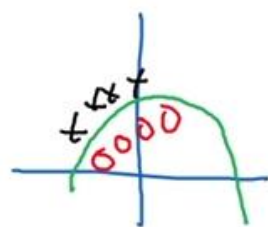
↑

$$z^{[L]} = w^{[L]} a^{[L-1]}$$

Ковариационный сдвиг



Cat Non-Cat
 $y = 1$ ✓ $y = 0$



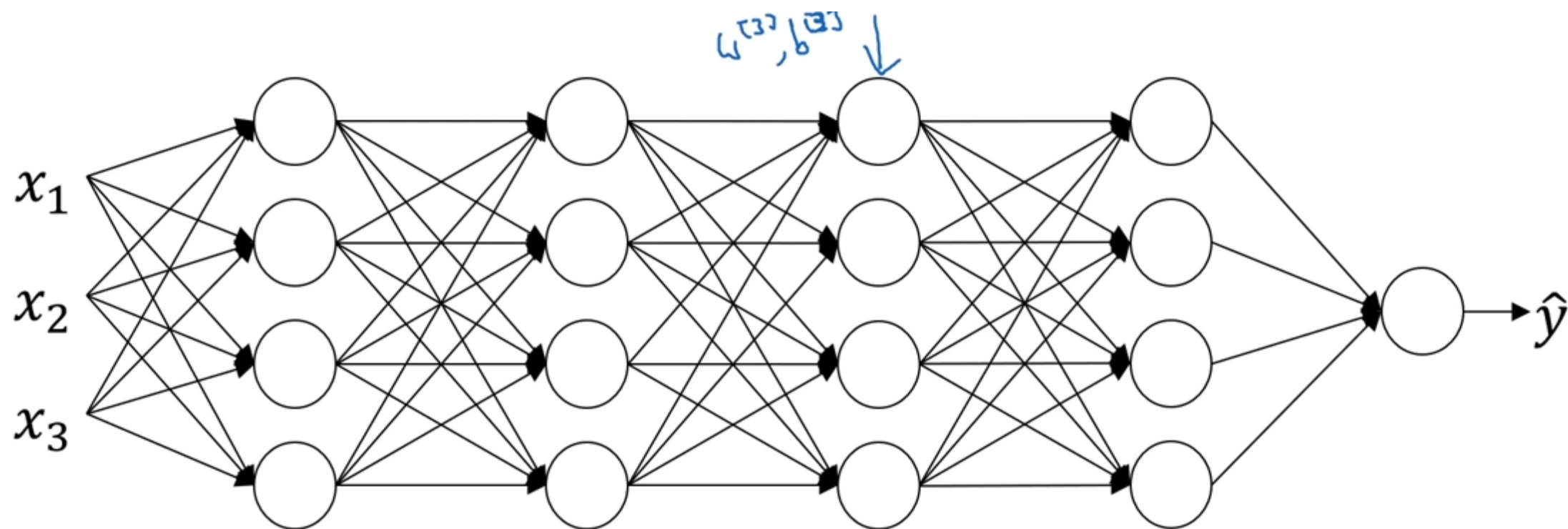
$y = 1$ ✓ $y = 0$



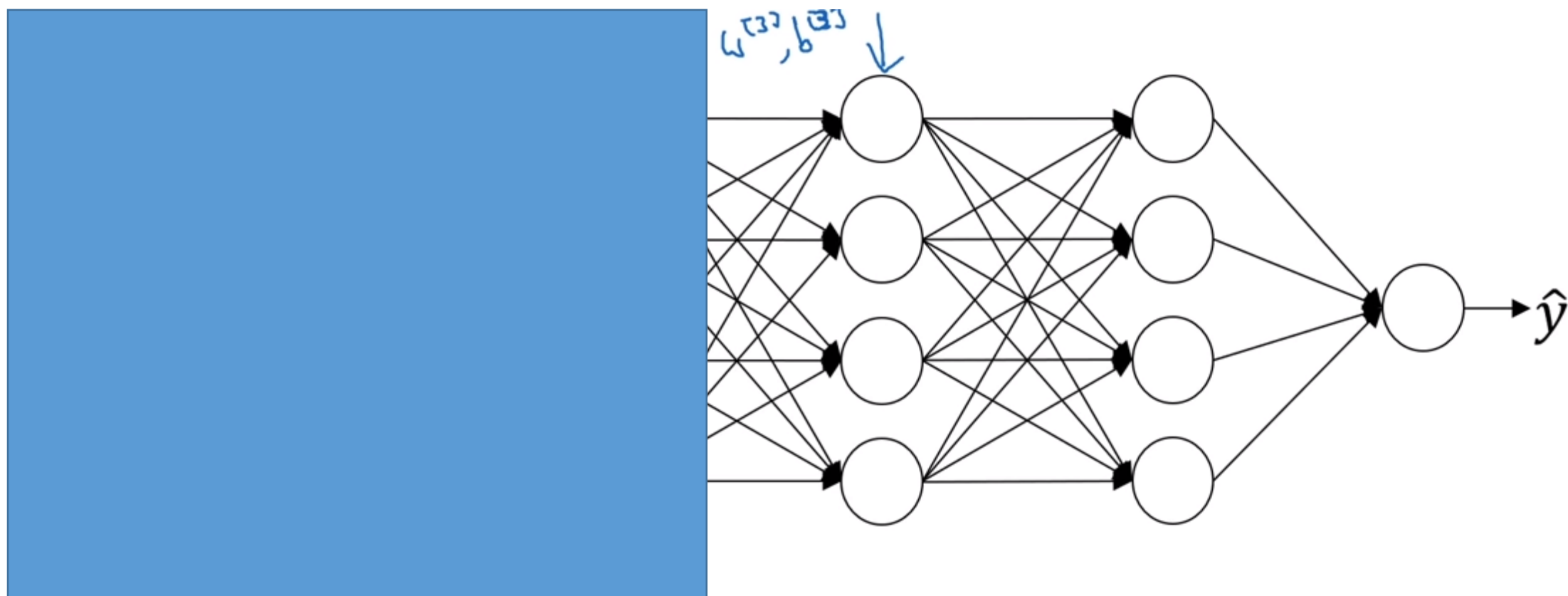
"Covariate shift"

$\underline{x} \rightarrow y$

Ковариационный сдвиг в НС



Ковариационный сдвиг в НС



VN в режиме тестирования

- Если тестируемся батчами

- $\mu = \frac{1}{m} \sum z^i$

- $\sigma^2 = \frac{1}{m} \sum (z_i - \mu)^2$

- $z_{norm}^i = \frac{z^i - \mu}{\sqrt{\sigma^2 + \epsilon}}$

- $\tilde{z}^i = \gamma z_{norm}^i + \beta$

В процессе тренировки считаем скользящее экспоненциальное среднее

$$\begin{aligned} s_t &= \theta s_{\{t-1\}} + (1 - \theta) \sigma_t^2 v_t \\ &= \theta v_{\{t-1\}} + (1 - \theta) \mu_t \end{aligned}$$

В режиме теста используем эти значения

$$\begin{aligned} z_{norm}^i &= \frac{z^i - v_t}{\sqrt{s_t + \epsilon}} \\ \tilde{z}^i &= \gamma z_{norm}^i + \beta \end{aligned}$$