

RL: Контекстный многорукий бандит

Многорукий бандит



Многорукий бандит: простая модель

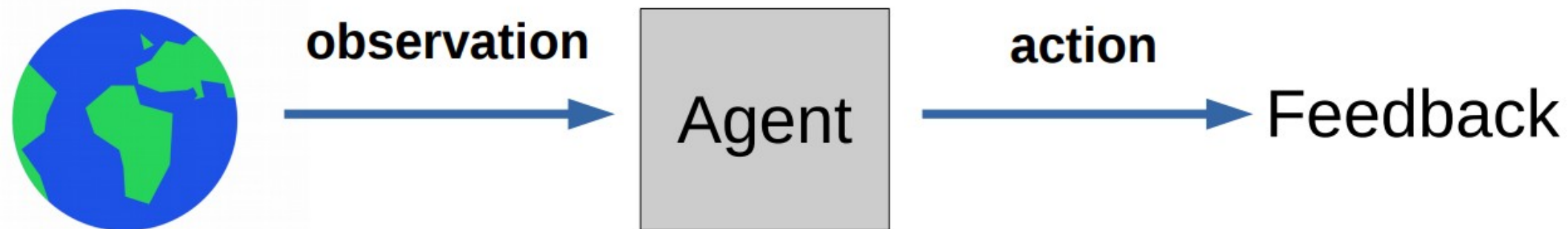
Простой случай: нет различных "состояний", только N действий



Exploration: выяснить, какое действие в целом лучше;
сделать как можно меньше плохих действий

Многорукий бандит: контекст

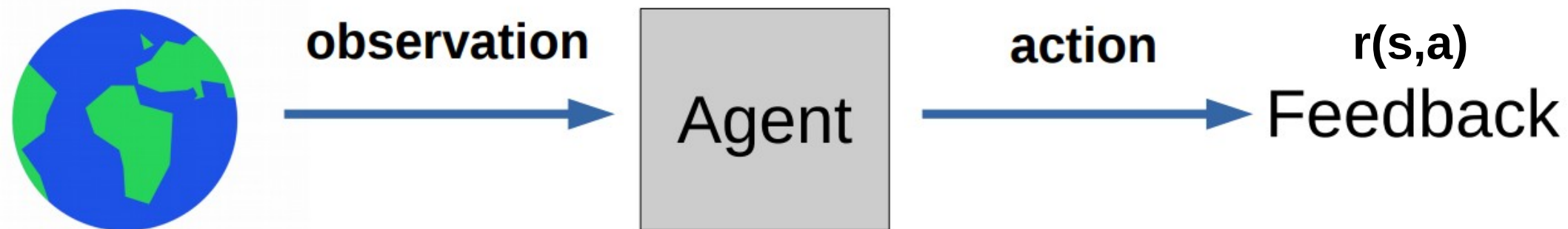
Упрощенный MDP с одним шагом



Почему бандиты: здесь проще объяснять математику,
формулы примерно на 50% короче
(обобщении MDP далее)

Что такое контекстный бандит

Упрощенный MDP с одним шагом



Примеры:

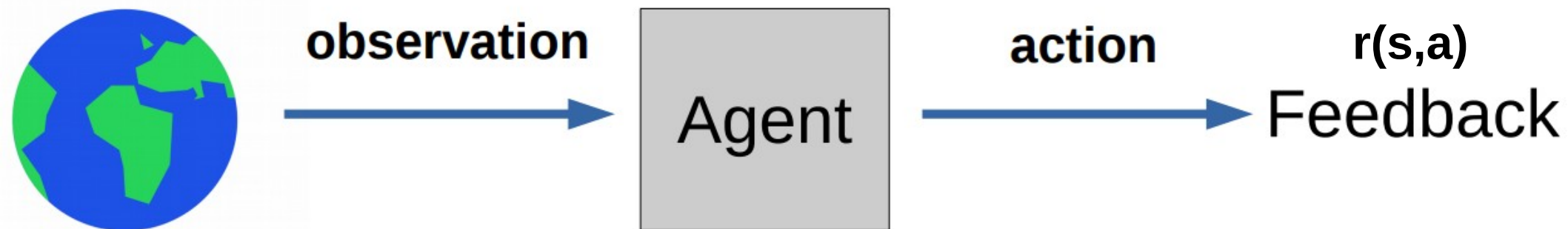
- Баннерная реклама
- Рекомендации
- Медицинское лечение

В основном это одношаговый MDP, где

- $G(s,a) = r(s,a)$
- $Q(s,a) = E r(s,a)$
- Все формулы на 50% короче

Что такое контекстный бандит

Упрощенный MDP с одним шагом



Примеры:

- Баннерная реклама
- Рекомендации
- Медицинское лечение

В основном это одношаговый MDP, где

- $G(s,a) = r(s,a)$
- $Q(s,a) = E r(s,a)$
- Все формулы на 50% короче

Что такое контекстный бандит

- Имеем набор автоматов (черных ящиков)
- Результат работы зависит от контекста



Примеры:

- Баннерная реклама: контекст - характеристик пользователя
- Рекомендации: контекст — выбор пользователя на предыдущих шагах
- Медицинское лечение: контекст — состояние здоровья

Контекстный бандит: решение «в лоб»

- Каждый контекст рассматривать как отдельный автомат

Пример:

- Баннерная реклама: контекст — пол (M,F) и OS (W,Mac)



Контекстный бандит: байесовский подход

$$w_0 + w_1 M + w_2 F + w_3 W + w_4 Mac + w_5 First + w_6 Second = r$$

$$w_0 \cdot 1 + w_1 x_1 + w_2 x_2 + w_3 x_3 + w_4 x_4 + w_5 x_5 + w_6 x_6 = r \Rightarrow w^T x = r$$

$$w = (x^T x)^{-1} x^T r$$

$$p(\theta | x) = \frac{p(x | \theta) p(\theta)}{p(x)}$$

$$p(w | y, x) = \frac{p(y | x, w) p(w)}{p(y)}$$

$$y = w^T x + \epsilon \quad \epsilon \sim N(\epsilon | 0, \sigma^2)$$

$$p(y | w, x) = N(y | w^T x, \sigma^2)$$

Контекстный бандит: байесовский подход

$$p(y|w, x) = N(y|w^T x, \sigma^2)$$

$$p(y|w, x) = \prod_i p(y_i|w, x_i) = \prod_i N(y_i|w^T x_i, \sigma^2) = \prod_i \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(y_i - w^T x_i)^2} \rightarrow \max$$

$$\ln p(y|w, x) = -n \ln \sqrt{2\pi}\sigma^2 - \sum_i \frac{1}{2\sigma^2} (y_i - w^T x_i)^2 \rightarrow \max$$

$$\ln p(y|w, x) = \text{const} - \sum_i \frac{1}{2\sigma^2} (y_i - w^T x_i)^2 \rightarrow \max$$

Контекстный бандит: байесовский подход

$$L = \sum_n \left(y_n - w^T x_n \right)^2 + \lambda \|w\|^2 \qquad p(w|D) = \frac{p(D|w)p(w)}{p(D)} \sim p(D|w)p(w)$$

$$p(w) = N(w|0, \alpha I) \qquad p(y|w, x) = N(y|w^T x, \sigma^2)$$

$$\ln p(w|D) = \frac{d}{2} \ln 2\pi - d \ln \sigma_0 - \frac{1}{2\sigma_0^2} w^T w + \text{const} - \frac{1}{2\sigma^2} \sum_n \left(y_n - w^T x_n \right)^2 = \dots$$
$$\frac{1}{\sigma^2} \sum_n \left(y_n - w^T x_n \right)^2 + \frac{1}{\sigma_0^2} w^T w \rightarrow \min$$

Контекстный бандит: байесовский подход

$$p(w|D) = \frac{p(D|w)p(w)}{p(D)} \sim p(D|w)p(w) \quad p(y|w, x) = N(y|w^T x, \sigma^2)$$

$$p(w) = N(w|\mu_0, \Sigma_0) = \frac{2}{(\sqrt{2\pi})^d \sqrt{\det \Sigma_0}} e^{-\frac{1}{2}(w - \mu_0)^T \Sigma_0^{-1} (w - \mu_0)}$$

$$\ln p(w|D) = \text{const} - \frac{1}{2}(w - \mu_0)^T \Sigma_0^{-1} (w - \mu_0) - \frac{1}{2\sigma^2} \sum_n (y_n - w^T x_n)^2 = \dots$$

$$\text{const} - \frac{1}{2}(w^T \Sigma_0^{-1} w - 2w^T \Sigma_0^{-1} \mu_0) - \frac{1}{2\sigma^2}(w^T X^T X w - 2w^T X^T y)$$

Контекстный бандит: байесовский подход

$$const - \frac{1}{2} \left(w^T \Sigma_0^{-1} w - 2 w^T \Sigma_0^{-1} \mu_0 \right) - \frac{1}{2 \sigma^2} \left(w^T X^T X w - 2 w^T X^T y \right) = \dots$$

$$const - \frac{1}{2} \left(w^T \left(\Sigma_0^{-1} + \frac{1}{\sigma^2} X^T X \right) w - 2 w^T \left(\Sigma_0^{-1} \mu_0 + \frac{1}{\sigma^2} X^T y \right) \right) = \dots$$

$$const - \frac{1}{2} (w - \mu_N)^T \left(\Sigma_0^{-1} + \frac{1}{\sigma^2} X^T X \right) (w - \mu_N)$$

$$p(w | D) = N(w | \mu_n, \Sigma_n)$$

$$\Sigma_n = \left(\Sigma_0^{-1} + \frac{1}{\sigma^2} X^T X \right)^{-1} \quad \mu_n = \Sigma_n \left(\Sigma_0^{-1} \mu_0 + \frac{1}{\sigma^2} X^T y \right)$$

Контекстный бандит: предсказание

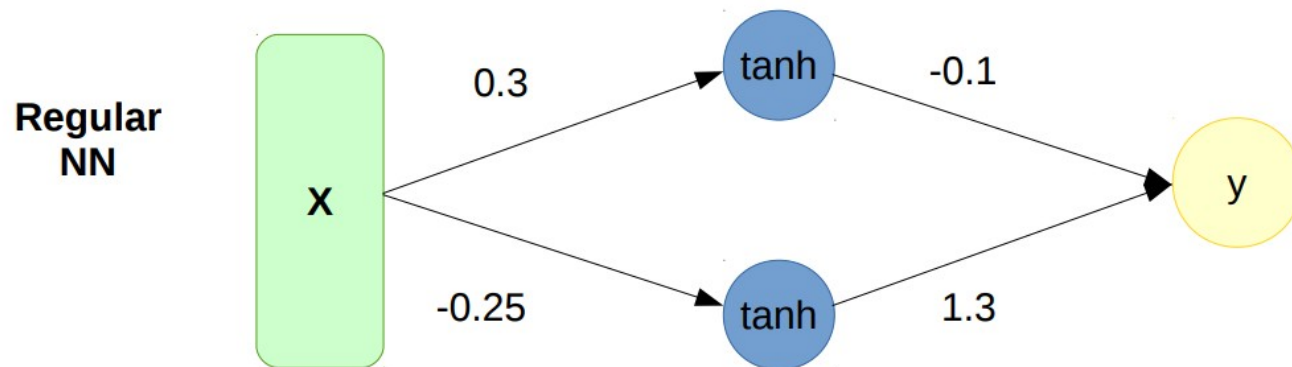
$$p(y|x, D) = \int p(y, w|x, D) dw = \int p(y|w, x) p(w|D) dw$$

$$p(y|x, D) = \int N(y|w^T x, \sigma^2) N(w|\mu_N, \Sigma_N) dw$$

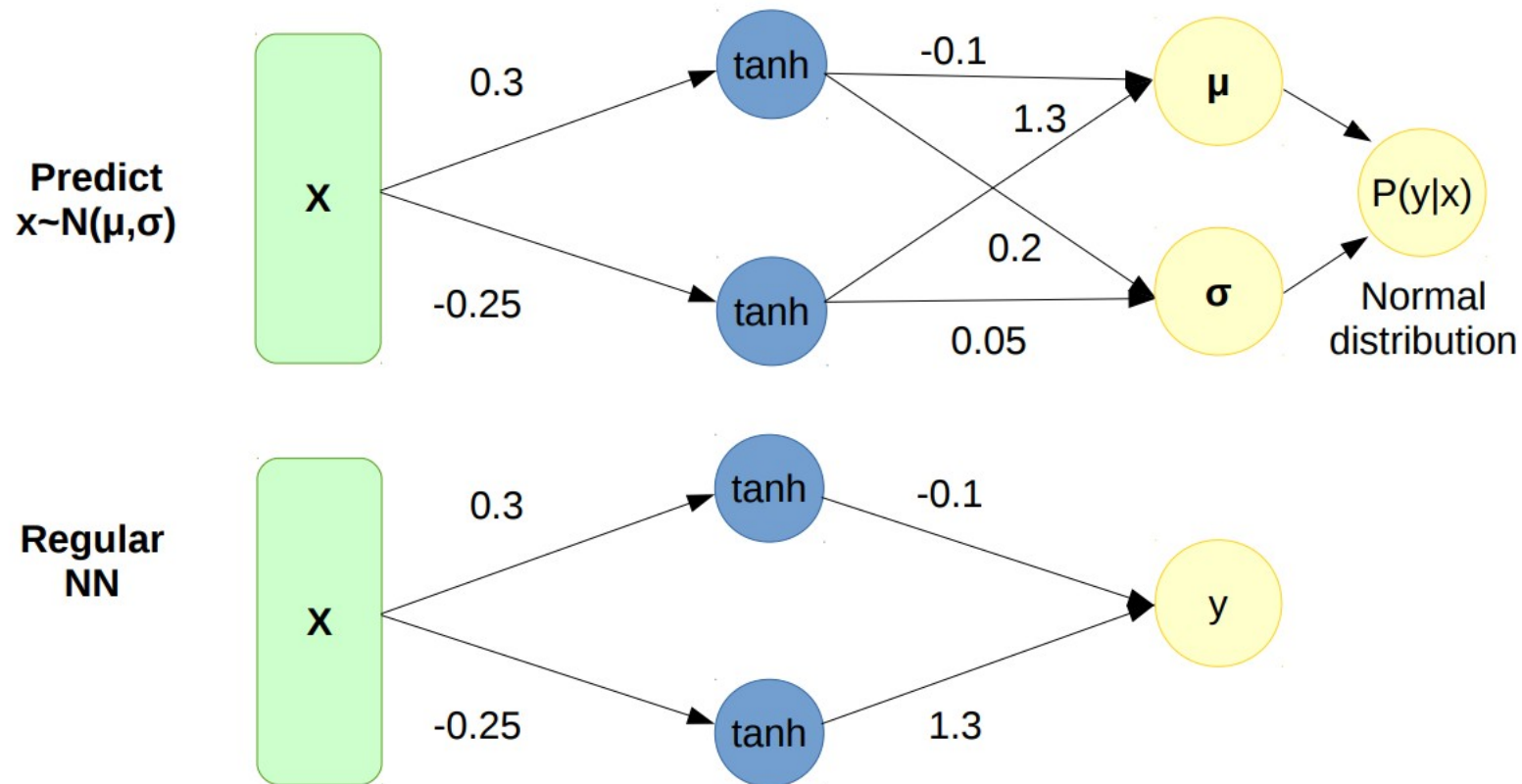
$$\sigma_{pred}^2 = \sigma^2 + x^T \Sigma_0^{-1} x$$

$$\mu_{pred} = \mu_N$$

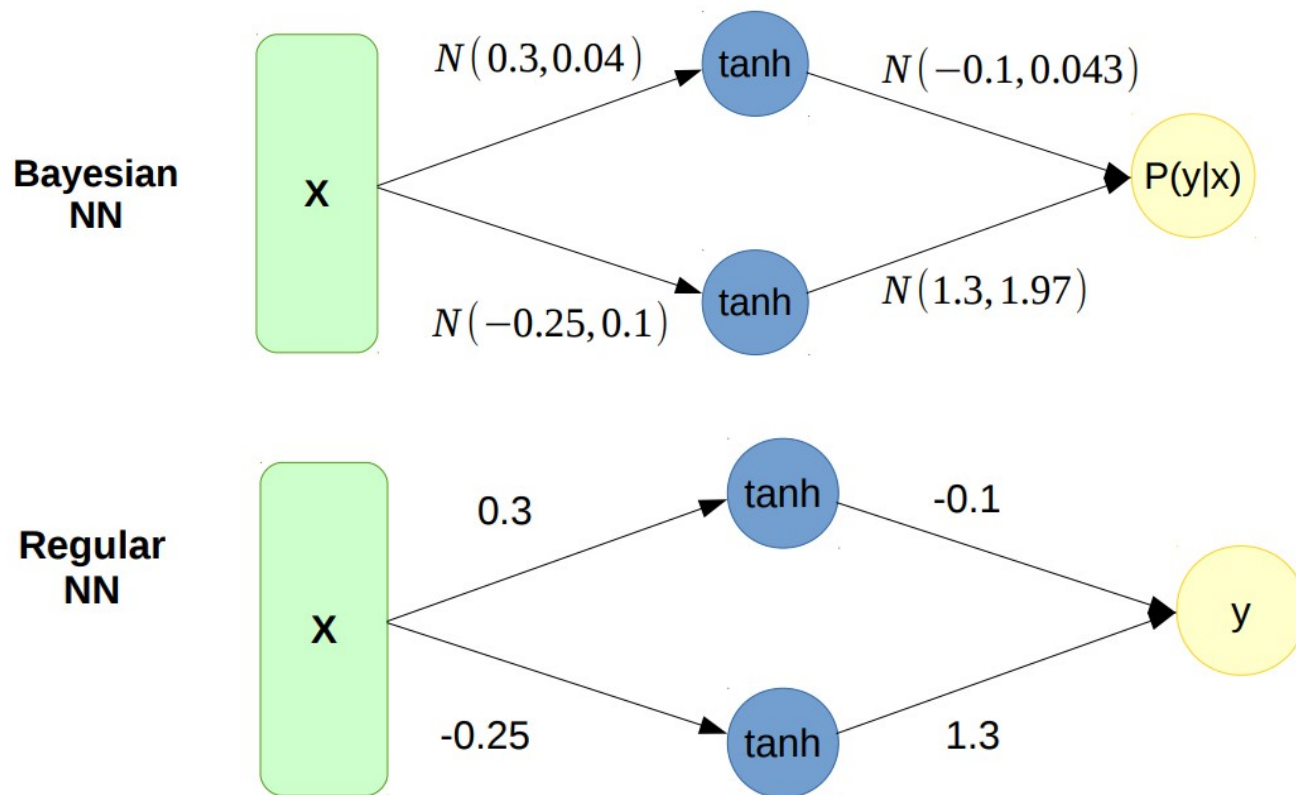
Параметрическая оценка



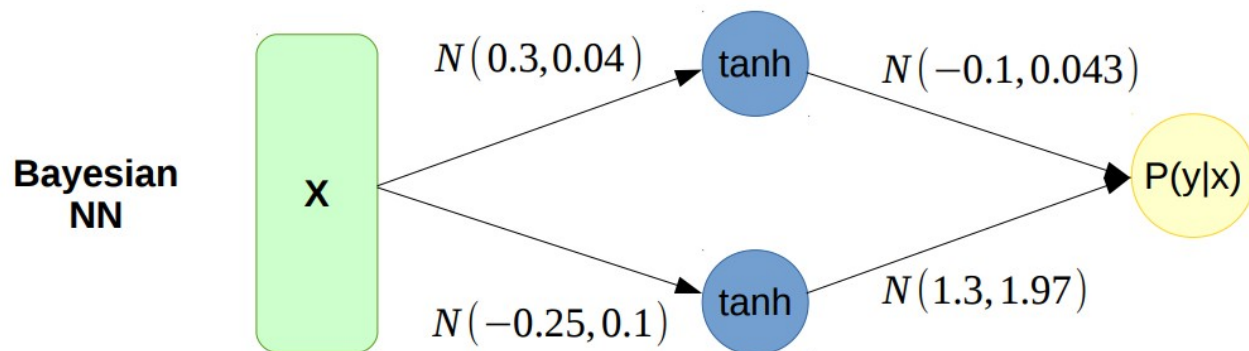
Параметрическая оценка



Байесовские нейронные сети



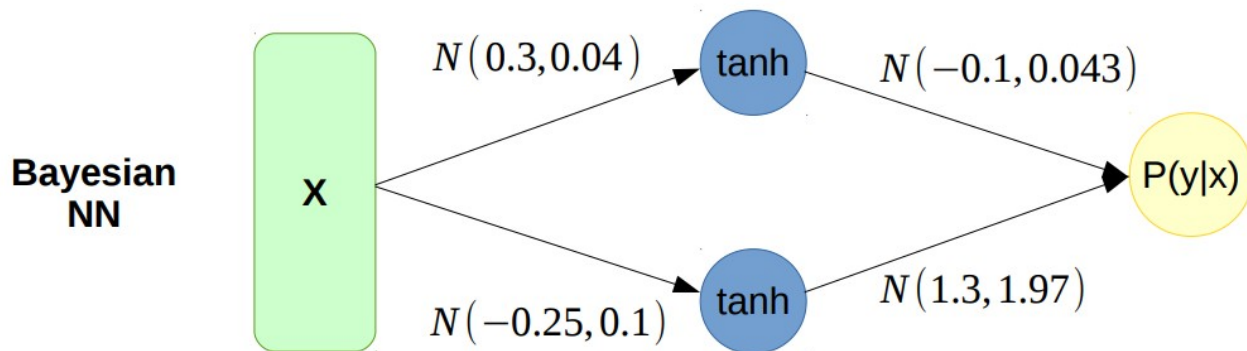
Байесовские нейронные сети



Идея:

- Никаких явных весов
- Поддерживать параметрическое распределение на весах
- Практика: полнофакторное нормальное или аналогичное

Байесовские нейронные сети



Идея:

- Никаких явных весов
- Inference: выборка из распределений весов, предсказание 1 точки
- Чтобы получить распределение, объединить K выборок (например, с помощью гистограммы)
- Да, это означает многократный прогон сети для одного X