Искусственный интеллект в науках о Земле

Михаил Криницкий

к.т.н.
Зав. лабораторией машинного обучения в науках о Земле МФТИ
с.н.с. Института океанологии РАН им. П.П. Ширшова

Линейная регрессия

$$y_i \sim \mathcal{N}(\mu(\theta, x_i), \sigma^2)$$

$$\eta_i = \theta \cdot x_i$$

$$\mu(\theta, x_i) = \boldsymbol{\eta_i}$$

Логистическая регрессия

$$y_i \sim \mathcal{B}\big(p(\theta, x_i)\big)$$

$$p(\theta, x_i) \propto \exp(\theta_1 \cdot x_i)$$

$$\eta_i^1 = \theta_1 \cdot x_i = \ln \frac{p_i}{1 - p_i}$$

$$p(\theta_1, x_i) = \mathbf{sigmoid}(\boldsymbol{\eta_i^1}) = \frac{1}{1 + \exp(-\eta_i^1)}$$

Мультиномиальная логистическая регрессия

$$y_{ik} \sim \mathcal{B}\big(p_k(\theta, x)\big)$$

$$p_k(\theta, x_i) \propto \exp(\theta_k \cdot x_i)$$

$$\eta_i^k = \theta_k \cdot x_i = \ln p_{ik} + C$$

$$p_k(\theta_1, x_i) = \mathbf{softmax}\big(\boldsymbol{\eta_i^k}, \{\boldsymbol{\eta_i^k}\}\big) = \frac{\exp \eta_i^k}{C^*}$$

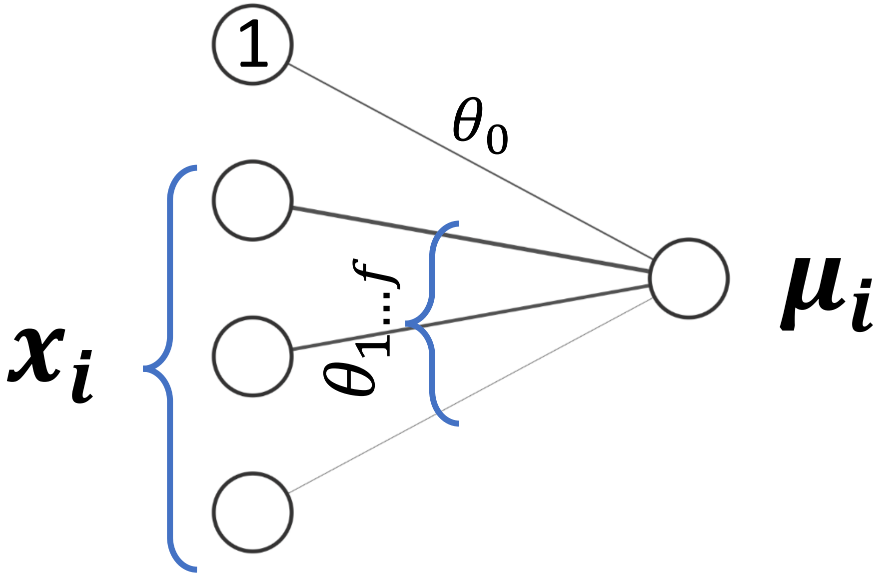
Альтернативно (некорректно, но легче понять):

Обобщенные линейные модели: модели, в которых мат.ожидание параметра распределения целевой переменной вычисляется как некоторая **функция** $\boldsymbol{g^{-1}}(\cdot)$ от линейной функции η признакового описания объектов (событий) x_i

$\boldsymbol{g}(\cdot)$ – т.н. функция связи (link function)

Диаграммы GLM

Линейная регрессия



$$\boldsymbol{y} \sim \boldsymbol{\mathcal{N}}(\boldsymbol{\mu}, \boldsymbol{\sigma^2})$$

$$\eta_i = \theta \cdot x_i$$

$$\mu(\theta, x_i) = \boldsymbol{\eta_i}$$

$$\mathcal{L}(\mathcal{T}, \theta) = C \sum_{\mathcal{T}} (y_i - \mu_i)^2$$

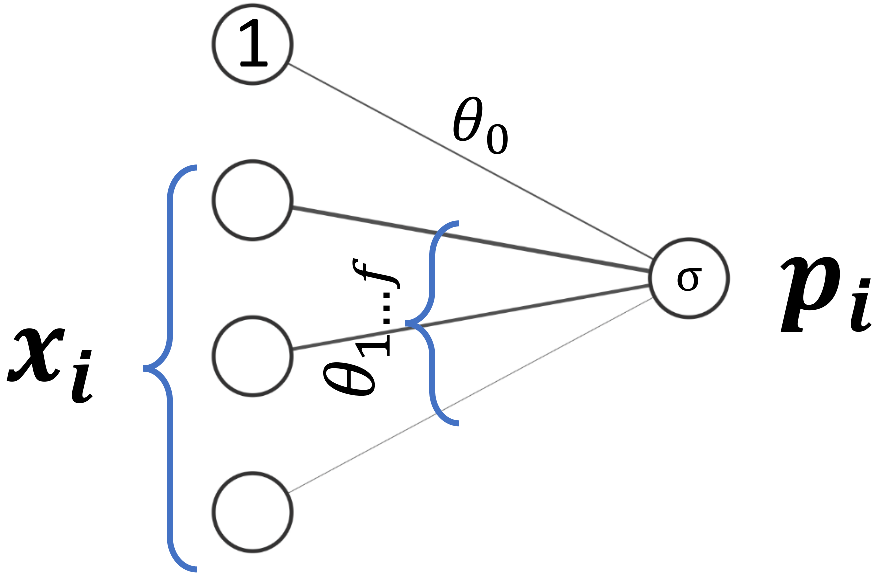
$$\theta = \operatorname*{argmin}_{\Theta}\big(\mathcal{L}(\mathcal{T}, \theta)\big)$$

$$\frac{\partial \mathcal{L}(\mathcal{T}, \theta)}{\partial \theta} = -2X^T Y + 2X^T X\theta$$

=> градиентная оптимизация

Диаграммы GLM

Логистическая регрессия



$$\boldsymbol{y_i} \sim \boldsymbol{\mathcal{B}}\big(\boldsymbol{p}(\boldsymbol{\theta}, \boldsymbol{x_i})\big)$$

$$\eta_i^1 = \theta_1 \cdot x_i$$

$$p(\theta_1, x_i) = \boldsymbol{\sigma}(\boldsymbol{\eta_i^1})$$

$$\mathcal{L}(\mathcal{T}, \theta) = -\sum_{\mathcal{T}} (y_i * \log p_i + (1 - y_i) * \log(1 - p_i))$$

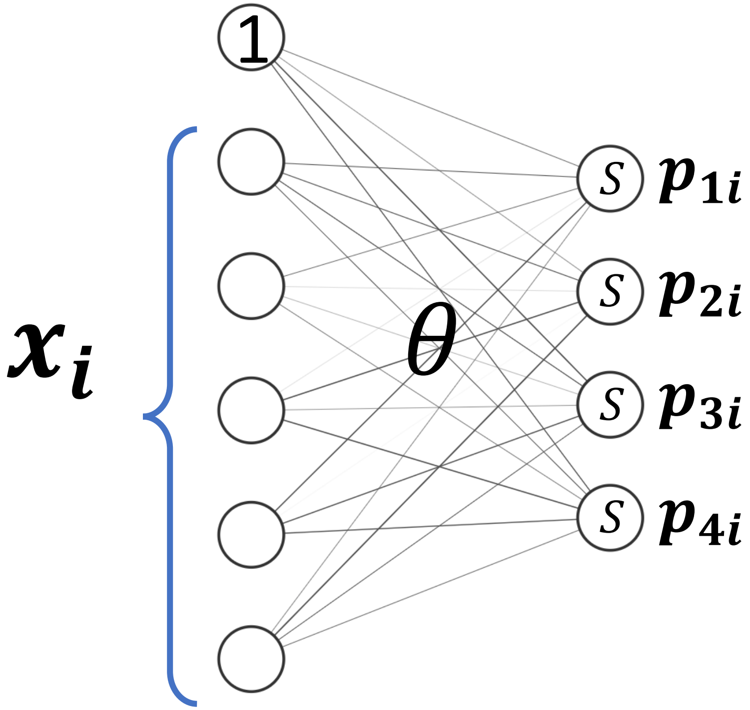
$$\theta = \underset{\Theta}{\operatorname{argmin}}\big(\mathcal{L}(\mathcal{T}, \theta)\big)$$

$$\frac{\partial \mathcal{L}(\mathcal{T}, \theta)}{\partial \theta} = -\sum_i \big(y_i - p(\theta, x_i)\big) x_i$$

=> градиентная оптимизация

Диаграммы GLM

Мультиномиальная
логистическая регрессия



$$y_{ik} \sim \mathcal{B}\big(p_k(\theta, x)\big)$$

$$\eta_i^k = \theta_k \cdot x_i = \ln p_{ik} + C$$

$$p_k(\theta_1, x_i) = \mathbf{softmax}\big(\boldsymbol{\eta_i^k}, \{\boldsymbol{\eta_i^k}\}\big) = \frac{\exp \eta_i^k}{C^*}$$

$$\mathcal{L}(\mathcal{T}, \theta) = -\sum_{\mathcal{T}} \sum_{k=1}^{K} ([y_i == k] * \log p_{ik})$$

$$\theta = \underset{\Theta}{\operatorname{argmin}}\big(\mathcal{L}(\mathcal{T}, \theta)\big)$$

$$\nabla_{\theta_k} \mathcal{L}(\mathcal{T}, \theta) = -\sum_{\mathcal{T}} ([y_i == k] - p_{ik}) x_i$$

=> градиентная оптимизация

Регрессия

$$y_i \sim \mathcal{N}(\mu(\theta, x_i), \sigma^2)$$

$$\eta_i = \theta \cdot \left[f_1\left(x_i^{(1)}\right), f_2\left(x_i^{(2)}\right), f_3\left(x_i^{(3)}\right) \dots f_f\left(x_i^{(f)}\right)\right]$$

$$\mu(\theta, x_i) = \eta_i$$

Бинарная классификация

$$y_i \sim \mathcal{B}\big(p(\theta, x_i)\big)$$

$$p(\theta, x_i) \propto \exp(\theta_1 \cdot x_i)$$

$$\eta_i^1 = \theta \cdot \left[f_1\left(x_i^{(1)}\right), f_2\left(x_i^{(2)}\right), f_3\left(x_i^{(3)}\right) \dots f_f\left(x_i^{(f)}\right)\right] = \ln \frac{p_i}{1 - p_i}$$

$$p(\theta_1, x_i) = \mathrm{sigmoid}(\eta_i^1) = \frac{1}{1 + \exp(-\eta_i^1)}$$

Мультиномиальная классификация

$$y_{ik} \sim \mathcal{B}\big(p_k(\theta, x)\big)$$

$$p_k(\theta, x_i) \propto \exp(\theta_k \cdot x_i)$$

$$\eta_i^k = \theta \cdot \left[f_1\left(x_i^{(1)}\right), f_2\left(x_i^{(2)}\right), f_3\left(x_i^{(3)}\right) \dots f_f\left(x_i^{(f)}\right)\right] = \ln p_{ik} + C$$

$$p(\theta_1, x_i) = \mathrm{softmax}\big(\eta_i^k, \{\eta_i^k\}\big) = \frac{\exp \eta_i^k}{C^*}$$

Обобщенные аддитивные модели – это обобщенные **линейные** модели, в которых некоторая функция $g(\cdot)$ мат.ожидания параметра распределения целевой переменной вычисляется как **линейная функция** некоторых других гладких функций (часто нелинейных, но необязательно) признакового описания объектов (событий)

Регрессия

$$y_i \sim \mathcal{N}(\mu(\theta, x_i), \sigma^2)$$

$$\eta_i = \theta \cdot \left[f_1\left(x_i^{(1)}\right), f_2\left(x_i^{(2)}\right), f_3\left(x_i^{(3)}\right) \dots f_f\left(x_i^{(f)}\right)\right]$$

$$\mu(\theta, x_i) = \boldsymbol{\eta_i}$$

Бинарная классификация

$$y_i \sim \mathcal{B}\big(p(\theta, x_i)\big)$$

$$p(\theta, x_i) \propto \exp(\theta_1 \cdot x_i)$$

$$\eta_i^1 = \theta \cdot \left[f_1\left(x_i^{(1)}\right), f_2\left(x_i^{(2)}\right), f_3\left(x_i^{(3)}\right) \dots f_f\left(x_i^{(f)}\right)\right] = \ln\frac{p_i}{1 - p_i}$$

$$p(\theta_1, x_i) = \mathbf{sigmoid}(\boldsymbol{\eta_i^1}) = \frac{1}{1 + \exp(-\eta_i^1)}$$

Мультиномиальная классификация

$$y_{ik} \sim \mathcal{B}\big(p_k(\theta, x)\big)$$

$$p_k(\theta, x_i) \propto \exp(\theta_k \cdot x_i)$$

$$\eta_i^k = \theta \cdot \left[f_1\left(x_i^{(1)}\right), f_2\left(x_i^{(2)}\right), f_3\left(x_i^{(3)}\right) \dots f_f\left(x_i^{(f)}\right)\right] = \ln p_{ik} + C$$

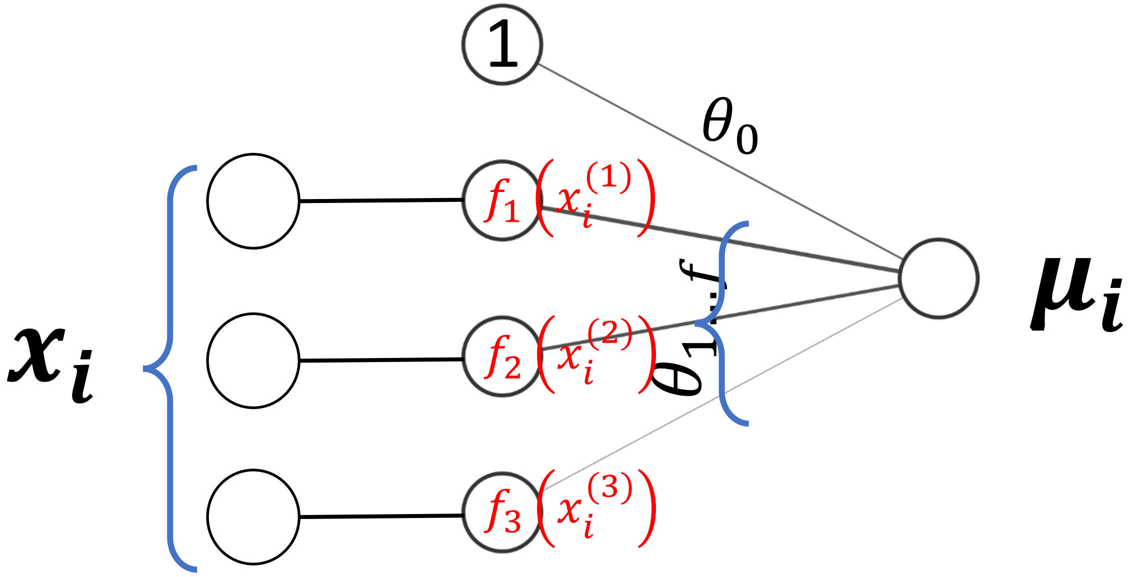
$$p(\theta_1, x_i) = \mathbf{softmax}\big(\boldsymbol{\eta_i^k}, \{\boldsymbol{\eta_i^k}\}\big) = \frac{\exp \eta_i^k}{C^*}$$

Альтернативно (некорректно, но легче понять):

Обобщенные аддитивные модели – это обобщенные линейные модели, в которых в качестве признаков объектов x_i выступают некоторые гладкие (часто нелинейные) функции этих признаков

Диаграммы GAM

Регрессия



$$\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$$

$$\eta_i = \theta \cdot \left[f_1\left(x_i^{(1)}\right), f_2\left(x_i^{(2)}\right), f_3\left(x_i^{(3)}\right) \ldots f_f\left(x_i^{(f)}\right)\right]$$

$$\mu(\theta, x_i) = \boldsymbol{\eta_i}$$

$$\mathcal{L}(\mathcal{T}, \theta) = C \sum_{\mathcal{T}} (y_i - \mu_i)^2$$

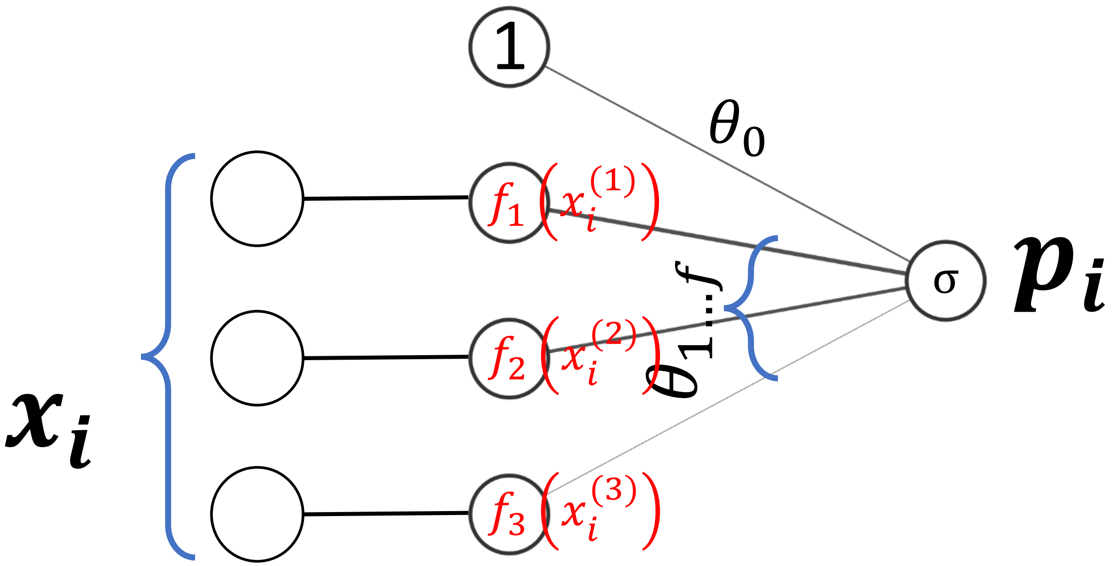
$$\theta = \underset{\Theta}{\operatorname{argmin}}\left(\mathcal{L}(\mathcal{T}, \theta)\right)$$

$$\frac{\partial \mathcal{L}(\mathcal{T}, \theta)}{\partial \theta} = ?$$

=> градиентная оптимизация

Диаграммы GAM

Бинарная классификация



$$\boldsymbol{y_i} \sim \boldsymbol{\mathcal{B}}\big(\boldsymbol{p}(\boldsymbol{\theta}, \boldsymbol{x_i})\big)$$

$$\eta_i^1 = \theta \cdot \left[f_1\left(x_i^{(1)}\right), f_2\left(x_i^{(2)}\right), f_3\left(x_i^{(3)}\right) \ldots f_f\left(x_i^{(f)}\right)\right]$$

$$p(\theta_1, x_i) = \boldsymbol{\sigma}\left(\boldsymbol{\eta_i^1}\right)$$

$$\mathcal{L}(\mathcal{T}, \theta) = -\sum_{\mathcal{T}} (y_i * \log p_i + (1 - y_i) * \log(1 - p_i))$$

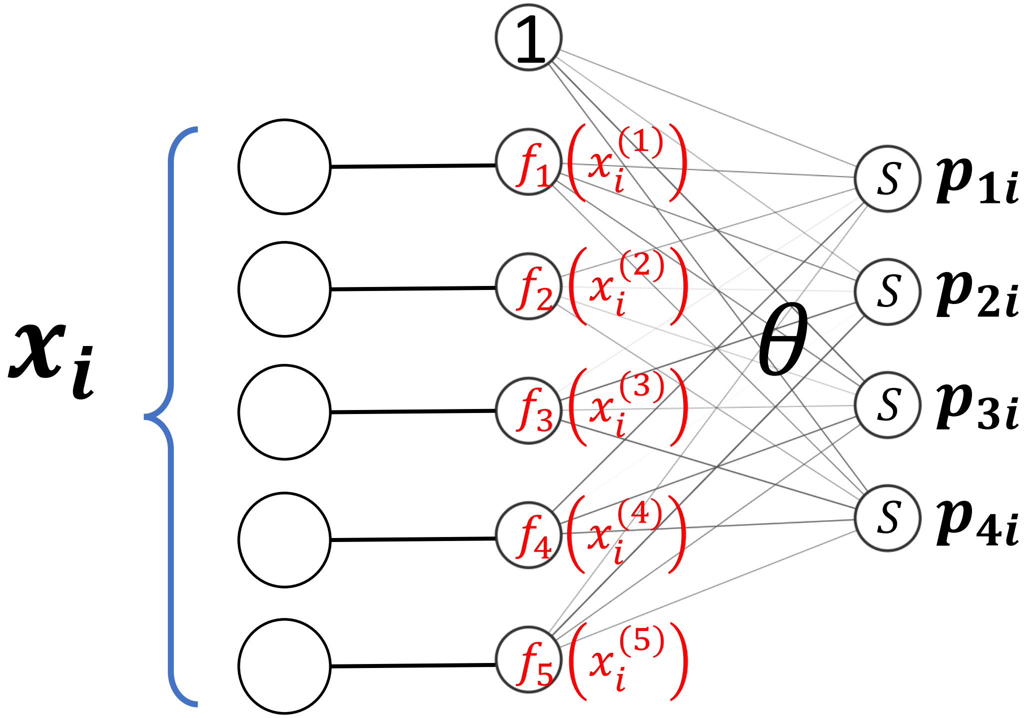
$$\theta = \underset{\Theta}{\operatorname{argmin}}\big(\mathcal{L}(\mathcal{T}, \theta)\big)$$

$$\frac{\partial \mathcal{L}(\mathcal{T}, \theta)}{\partial \theta} = ?$$

=> градиентная оптимизация

Диаграммы GAM

Мультиномиальная классификация



$$y_{ik} \sim \mathcal{B}\big(p_k(\theta, x)\big)$$

$$\eta_i^k = \theta \cdot \left[f_1\left(x_i^{(1)}\right), f_2\left(x_i^{(2)}\right), f_3\left(x_i^{(3)}\right) \dots f_f\left(x_i^{(f)}\right)\right]$$

$$\ln p_{ik} + C = \eta_i^k$$

$$p_k(\theta_1, x_i) = \mathbf{softmax}\left(\boldsymbol{\eta}_i^k, \{\boldsymbol{\eta}_i^k\}\right) = \frac{\exp \eta_i^k}{C^*}$$

$$\mathcal{L}(\mathcal{T}, \theta) = -\sum_{\mathcal{T}} \sum_{k=1}^{K} \left([y_i == k] * \log p_{ik}\right)$$

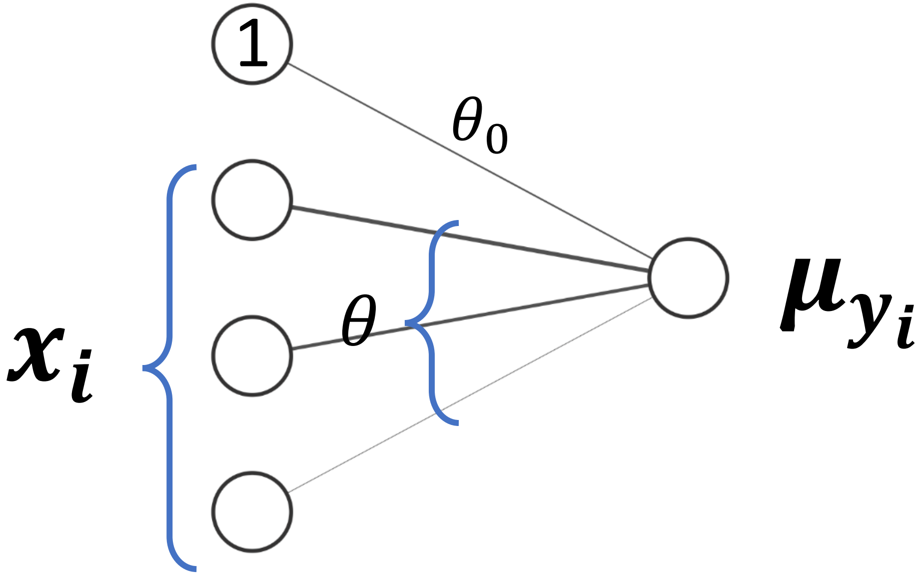
$$\theta = \underset{\Theta}{\operatorname{argmin}}\big(\mathcal{L}(\mathcal{T}, \theta)\big)$$

$$\nabla_{\theta_k} \mathcal{L}(\mathcal{T}, \theta) = ?$$

=> градиентная оптимизация

Почти всегда в случае задачи регрессии: $\boldsymbol{y} \sim \boldsymbol{\mathcal{N}}(\boldsymbol{\mu},\ \boldsymbol{\sigma^2})$

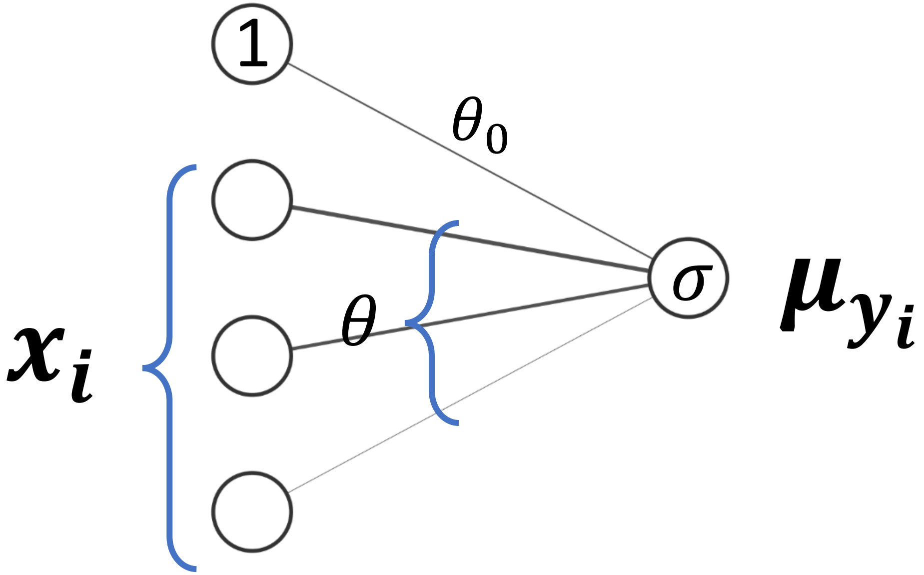
ЛР: $\mu_{y_i} = \theta \cdot x_i + \theta_0$



Почти всегда в случае задачи регрессии: $\boldsymbol{y} \sim \boldsymbol{\mathcal{N}}(\boldsymbol{\mu},\ \boldsymbol{\sigma}^2)$

ЛР: $\mu_{y_i} = \theta \cdot x_i + \theta_0$

GLM: $\mu_{y_i} = \sigma(\theta \cdot x_i + \theta_0)$

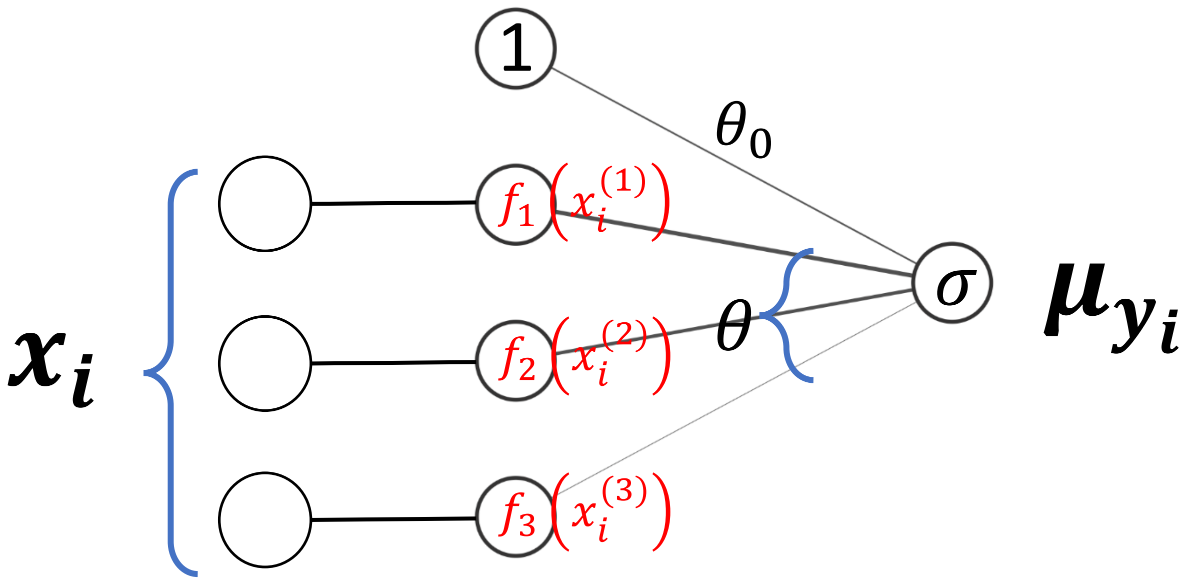


Почти всегда в случае задачи регрессии: $\boldsymbol{y} \sim \boldsymbol{\mathcal{N}}(\boldsymbol{\mu},\ \boldsymbol{\sigma}^2)$

ЛР: $\mu_{y_i} = \theta \cdot x_i + \theta_0$

GLM: $\mu_{y_i} = \sigma(\theta \cdot x_i + \theta_0)$

GAM: $\mu_{y_i} = \sigma\left(\theta \cdot \left[f_1\left(x_i^{(1)}\right), f_2\left(x_i^{(2)}\right), f_3\left(x_i^{(3)}\right) \dots f_f\left(x_i^{(f)}\right)\right] + \theta_0\right)$



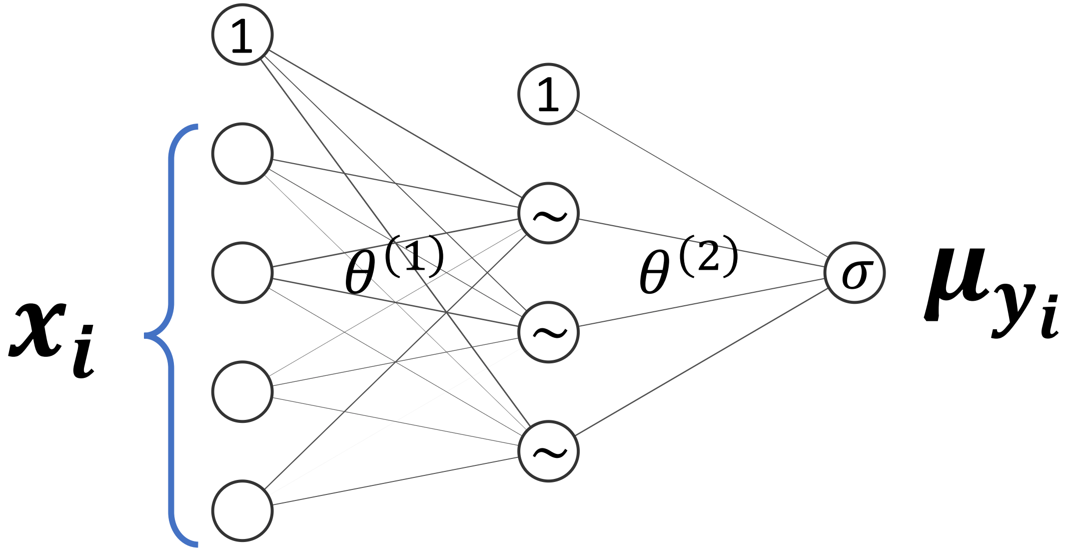
Почти всегда в случае задачи регрессии: $\boldsymbol{y} \sim \boldsymbol{\mathcal{N}}(\boldsymbol{\mu},\ \boldsymbol{\sigma}^2)$

ЛР: $$\mu_{y_i} = \theta \cdot x_i + \theta_0$$

GLM: $$\mu_{y_i} = \phi(\theta \cdot x_i + \theta_0)$$

GAM: $$\mu_{y_i} = \sigma\left(\theta \cdot \left[f_1\left(x_i^{(1)}\right), f_2\left(x_i^{(2)}\right), f_3\left(x_i^{(3)}\right) \dots f_f\left(x_i^{(f)}\right)\right] + \theta_0\right)$$

ИНС: $$\mu_{y_i} = \sigma\left(\theta_0^{(2)} + \theta^{(2)} \cdot \phi\left(\theta_0^{(1)} + \theta^{(1)} \cdot x_i\right)\right)$$



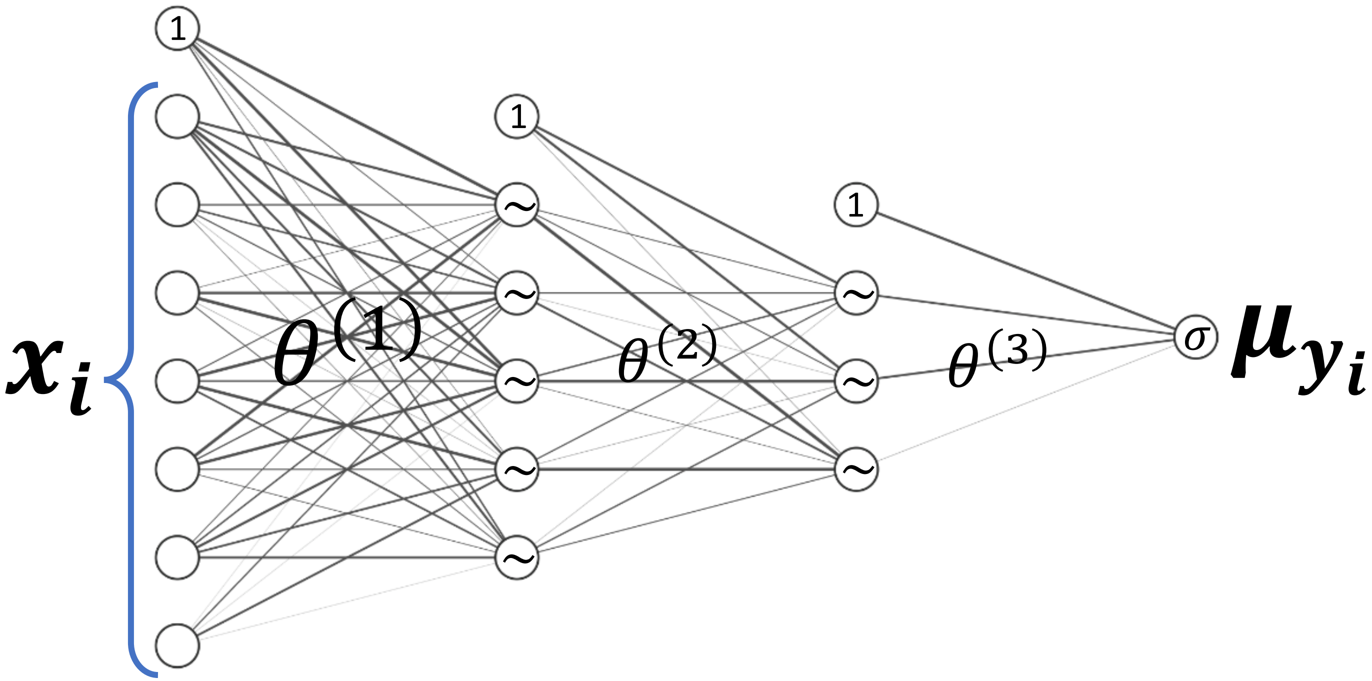
Почти всегда в случае задачи регрессии: $\boldsymbol{y} \sim \boldsymbol{\mathcal{N}}(\boldsymbol{\mu},\ \boldsymbol{\sigma}^2)$

ЛР: $\mu_{y_i} = \theta \cdot x_i + \theta_0$

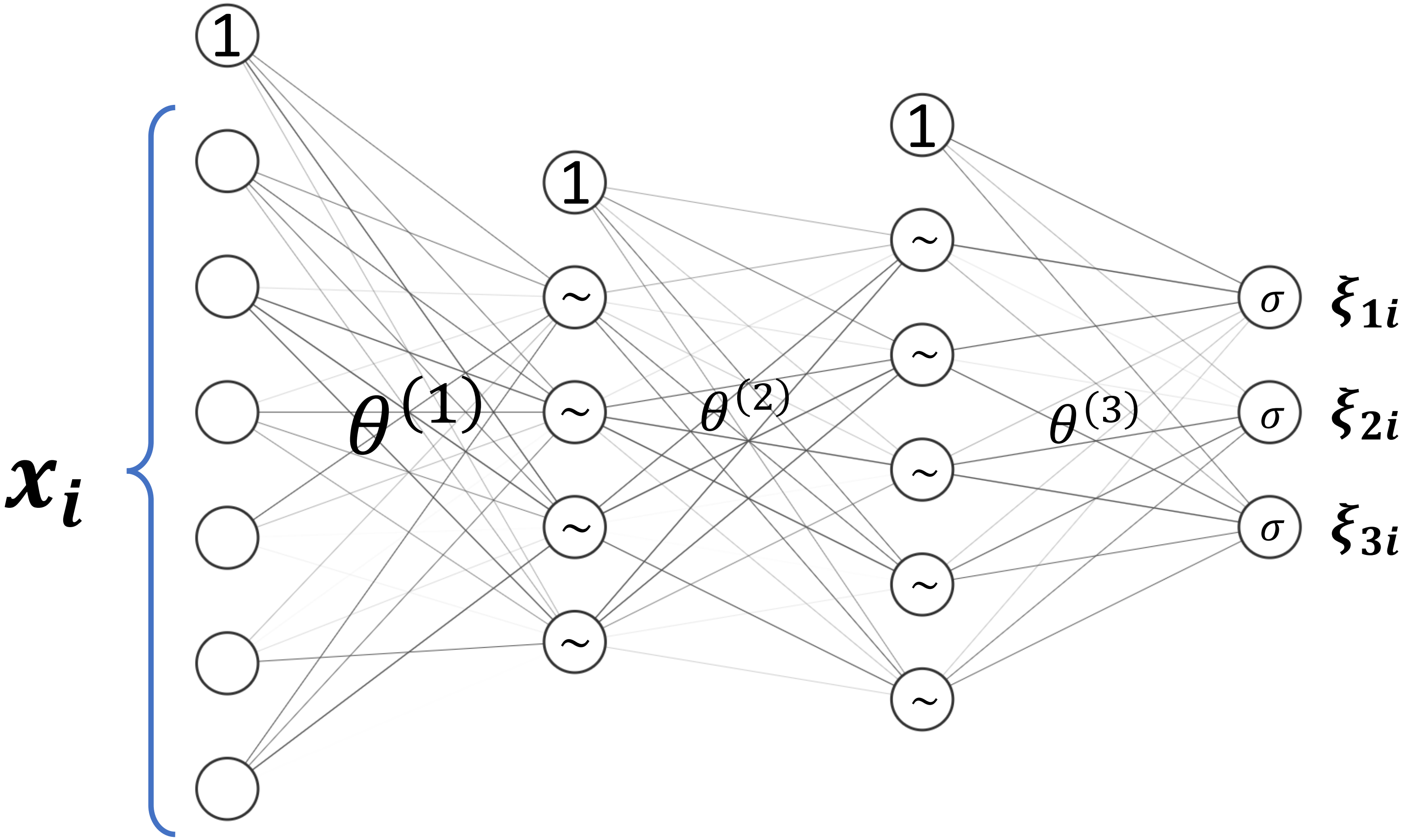
GLM: $\mu_{y_i} = \phi(\theta \cdot x_i + \theta_0)$

GAM: $\mu_{y_i} = \sigma\left(\theta \cdot \left[f_1\left(x_i^{(1)}\right), f_2\left(x_i^{(2)}\right), f_3\left(x_i^{(3)}\right) \dots f_f\left(x_i^{(f)}\right)\right] + \theta_0\right)$

ИНС: $\mu_{y_i} = \sigma\left(\theta_0^{(3)} + \theta^{(3)} \cdot \phi\left(\theta_0^{(2)} + \theta^{(2)} \cdot \phi\left(\theta_0^{(1)} + \theta^{(1)} \cdot x_i\right)\right)\right)$



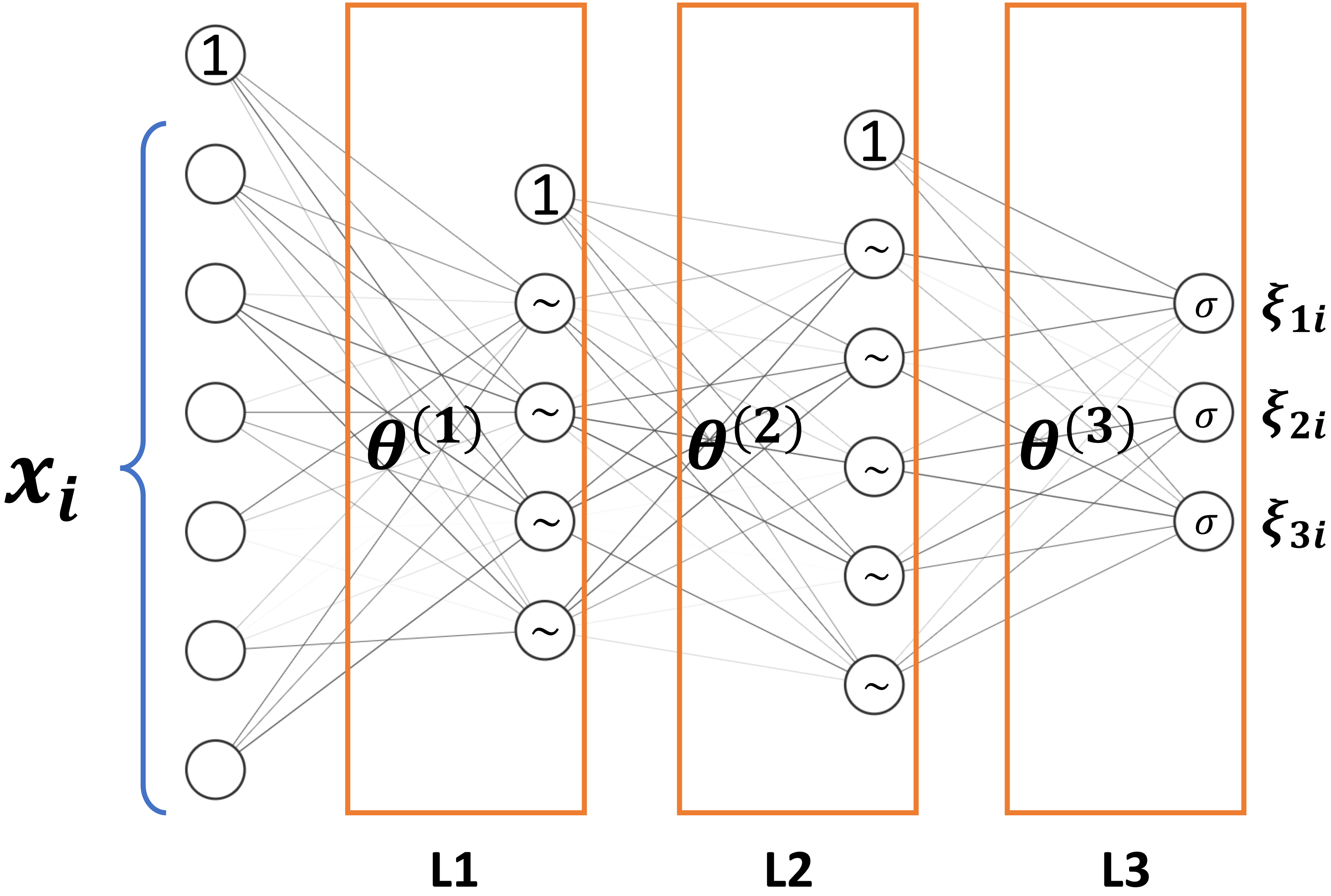
Искусственные нейронные сети



Вид ИНС: многослойный перцептрон (multilayer perceptron, MLP)
Feedforward NN (FNN, сеть прямого распространения)
полносвязная ИНС (Fully-connected NN, FCNN)

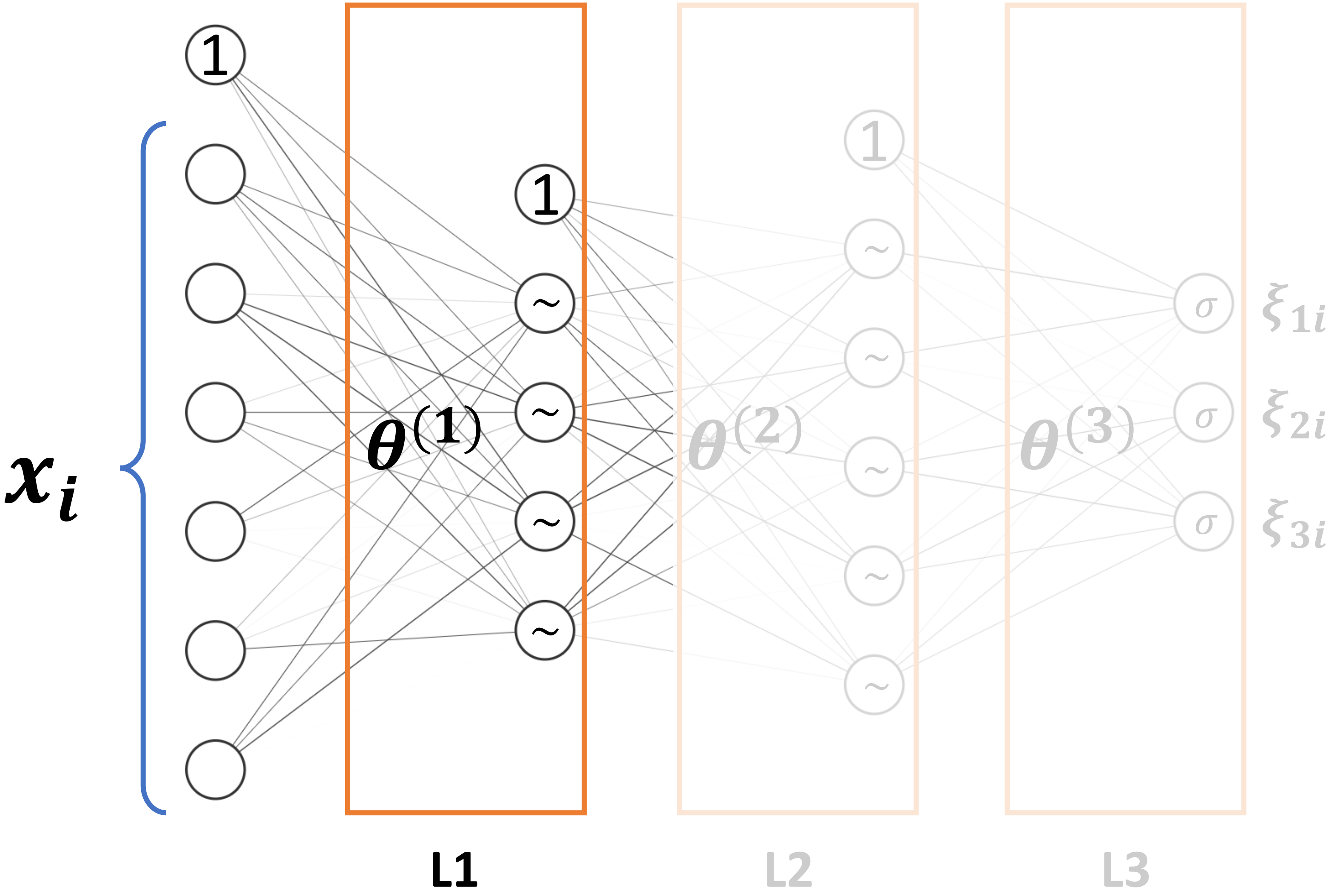
MLP

$$\xi_i = \sigma\left(\theta_0^{(3)} + \theta^{(3)} \cdot \phi\left(\theta_0^{(2)} + \theta^{(2)} \cdot \phi\left(\theta_0^{(1)} + \theta^{(1)} \cdot x_i\right)\right)\right)$$



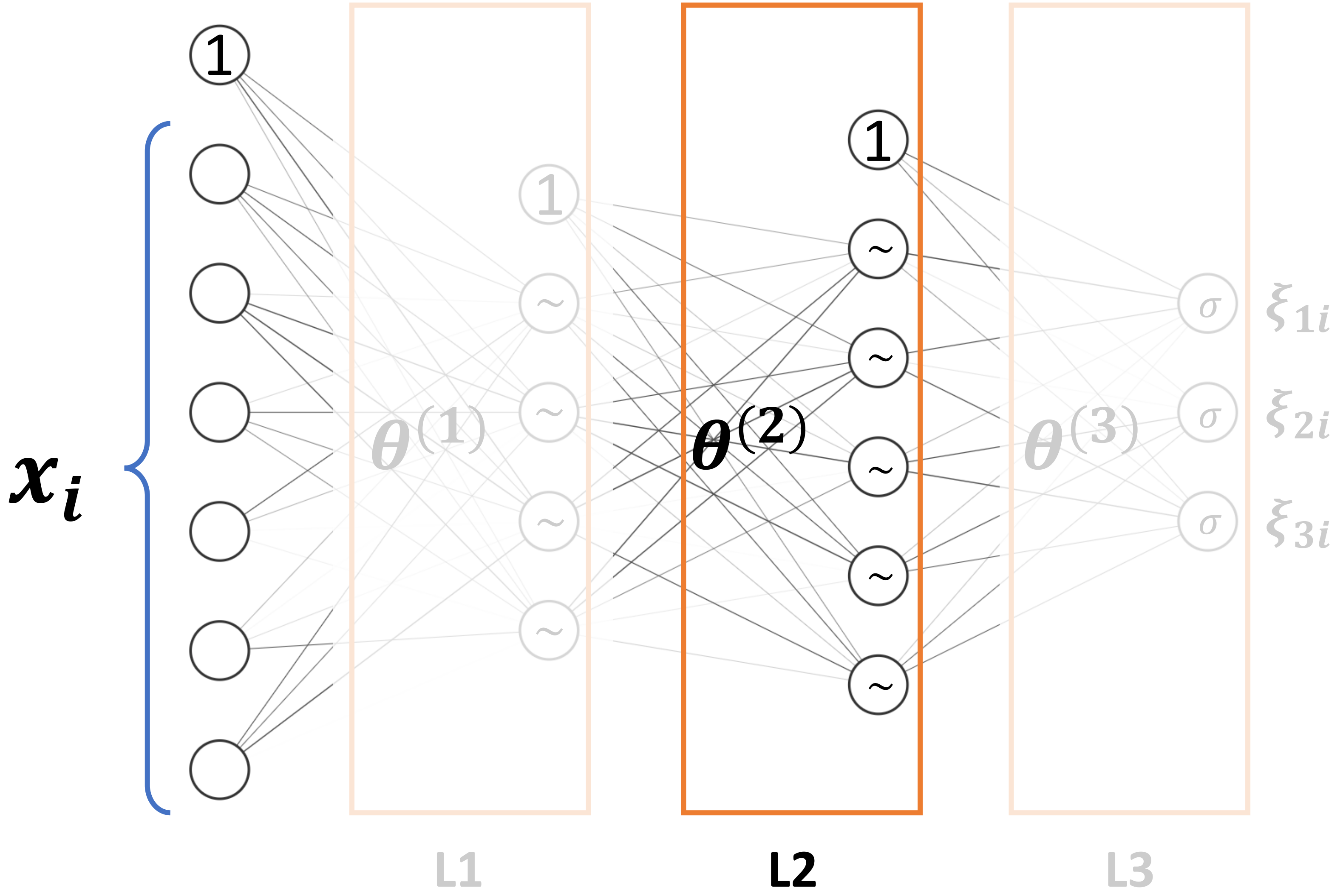
MLP

$$\xi_i = \sigma\left(\theta_0^{(3)} + \theta^{(3)} \cdot \phi\left(\theta_0^{(2)} + \theta^{(2)} \cdot \boxed{\phi\left(\theta_0^{(1)} + \theta^{(1)} \cdot x_i\right)}\right)\right)$$



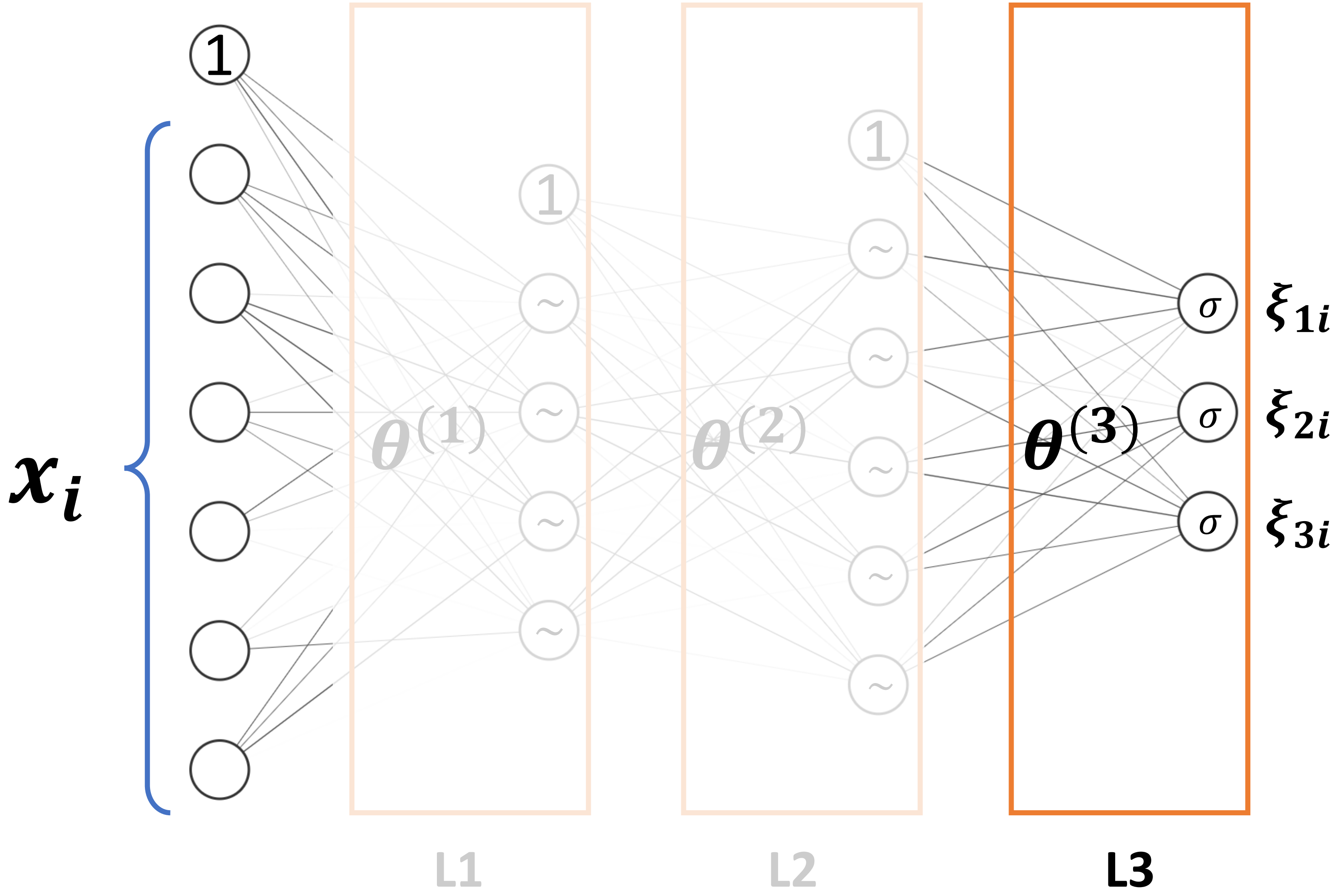
MLP

$$\xi_i = \sigma\left(\theta_0^{(3)} + \theta^{(3)} \cdot \boxed{\phi\left(\theta_0^{(2)} + \theta^{(2)} \cdot \phi\left(\theta_0^{(1)} + \theta^{(1)} \cdot x_i\right)\right)}\right)$$



MLP

$$\xi_i = \boxed{\sigma\left(\theta_0^{(3)} + \theta^{(3)} \cdot \phi\left(\theta_0^{(2)} + \theta^{(2)} \cdot \phi\left(\theta_0^{(1)} + \theta^{(1)} \cdot x_i\right)\right)\right)}$$



Обучение MLP

$$\xi_i = \boxed{\sigma\left(\theta_0^{(3)} + \theta^{(3)} \cdot \phi\left(\theta_0^{(2)} + \theta^{(2)} \cdot \phi\left(\theta_0^{(1)} + \theta^{(1)} \cdot x_i\right)\right)\right)}$$

Регрессия

$$\boldsymbol{y \sim \mathcal{N}(\mu, \sigma^2)}$$

$$\mu(\theta, x_i) = \boldsymbol{\xi_i}$$

$$\mathcal{L}(\mathcal{T}, \theta) = \sum_{\mathcal{T}} (y_i - \boldsymbol{\xi_i})^2$$

$$\theta = \underset{\Theta}{\operatorname{argmin}}\big(\mathcal{L}(\mathcal{T}, \theta)\big)$$

$$\frac{\partial \mathcal{L}(\mathcal{T}, \theta)}{\partial \theta} = \cdots$$

=> градиентная оптимизация

Бинарная классификация

$$\boldsymbol{y_i \sim \mathcal{B}\big(p(\theta, x_i)\big)}$$

$$p(\theta_1, x_i) = \boldsymbol{\xi_i}$$

$$\mathcal{L}(\mathcal{T}, \theta) = -\sum_{\mathcal{T}} (y_i * \log \boldsymbol{\xi_i} + (1 - y_i) * \log(1 - \boldsymbol{\xi_i}))$$

$$\theta = \underset{\Theta}{\operatorname{argmin}}\big(\mathcal{L}(\mathcal{T}, \theta)\big)$$

$$\frac{\partial \mathcal{L}(\mathcal{T}, \theta)}{\partial \theta} = \cdots$$

=> градиентная оптимизация

Мультиномиальная классификация

$$y_{ik} \sim \mathcal{B}\big(p_k(\theta, x)\big)$$

$$p_k(\theta_1, x_i) = \boldsymbol{\xi_{ik}}$$

$$\mathcal{L}(\mathcal{T}, \theta) = -\sum_{\mathcal{T}} \sum_{k=1}^{K} ([y_i == k] * \log \boldsymbol{\xi_{ik}})$$

$$\theta = \underset{\Theta}{\operatorname{argmin}}\big(\mathcal{L}(\mathcal{T}, \theta)\big)$$

$$\nabla_{\theta_k} \mathcal{L}(\mathcal{T}, \theta) = \ldots$$

=> градиентная оптимизация