Applied Parallel Computing
parallel-computing.pro

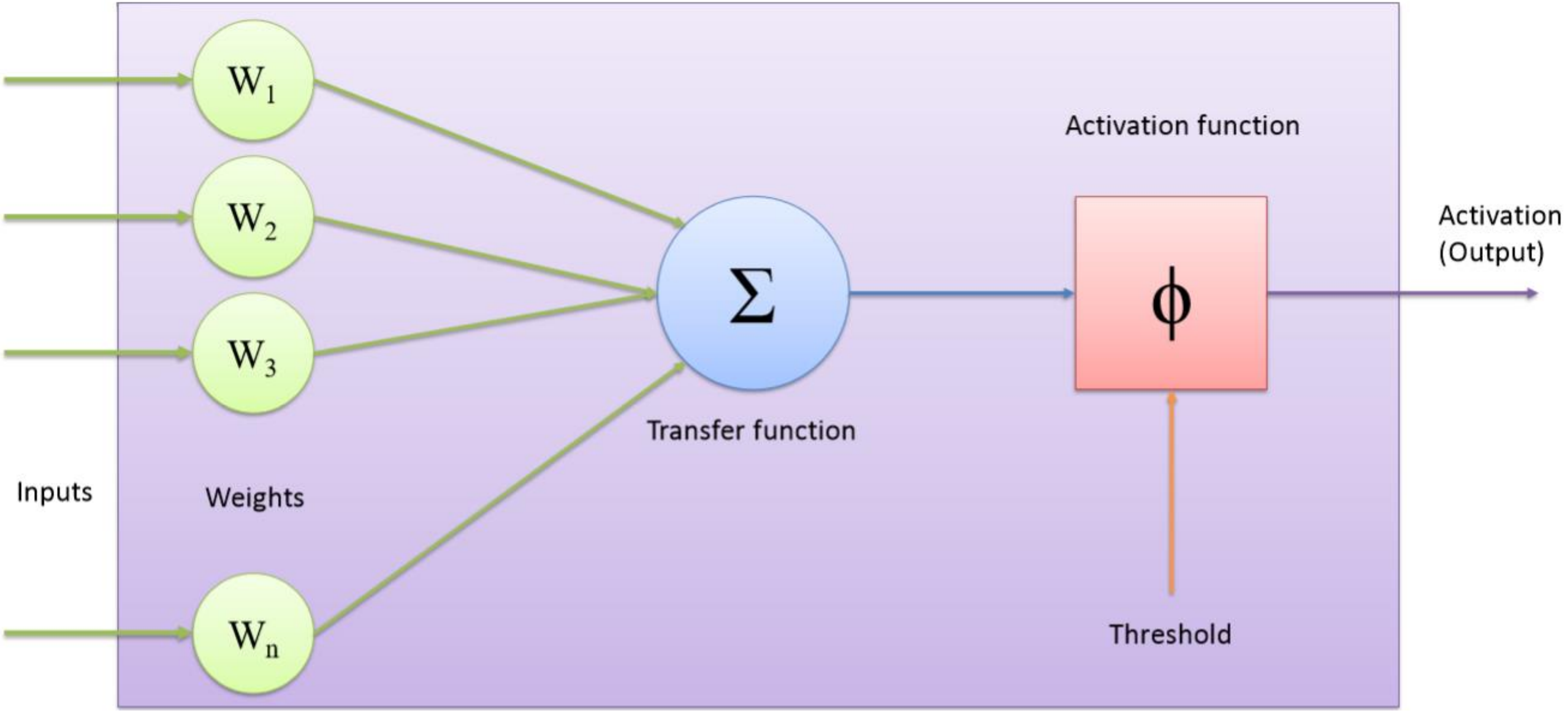
Обучение нейросетей

К.т.н. Алексей Ивахненко

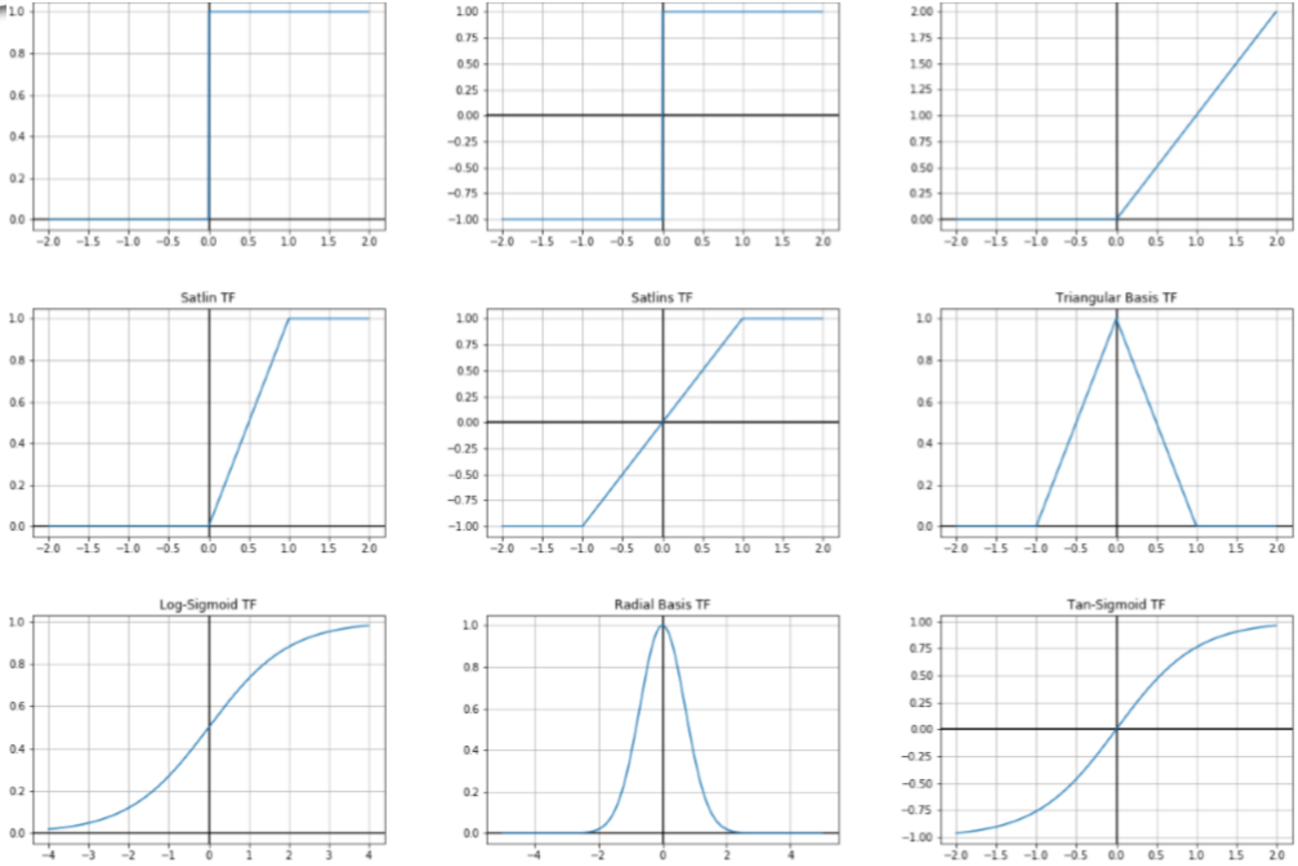
План

- Обучение нейросетей
  - С учителем
  - Без учителя
- Функции активации и производные
- Обратное распространение ошибки
- Метод градиентного спуска

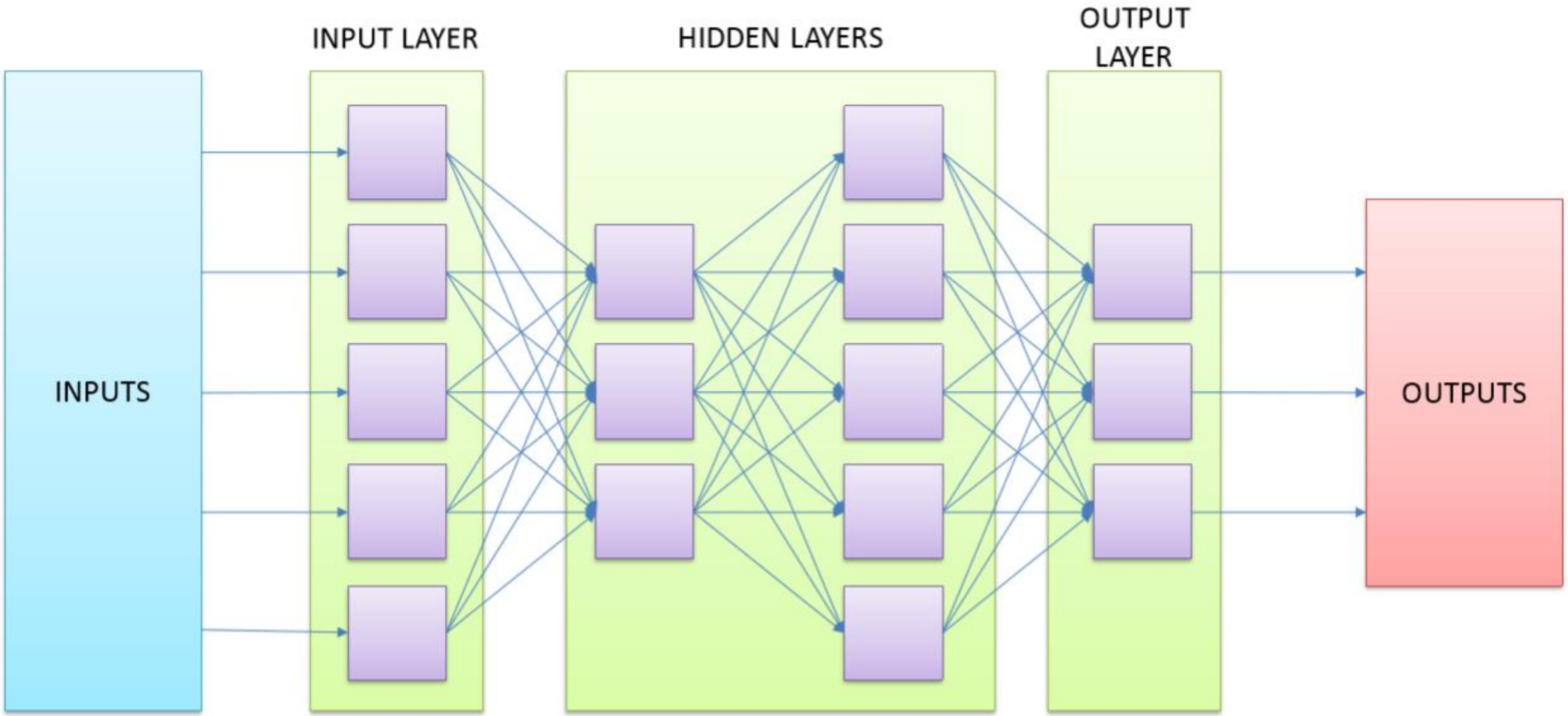
Искусственный нейрон



Функции активаций

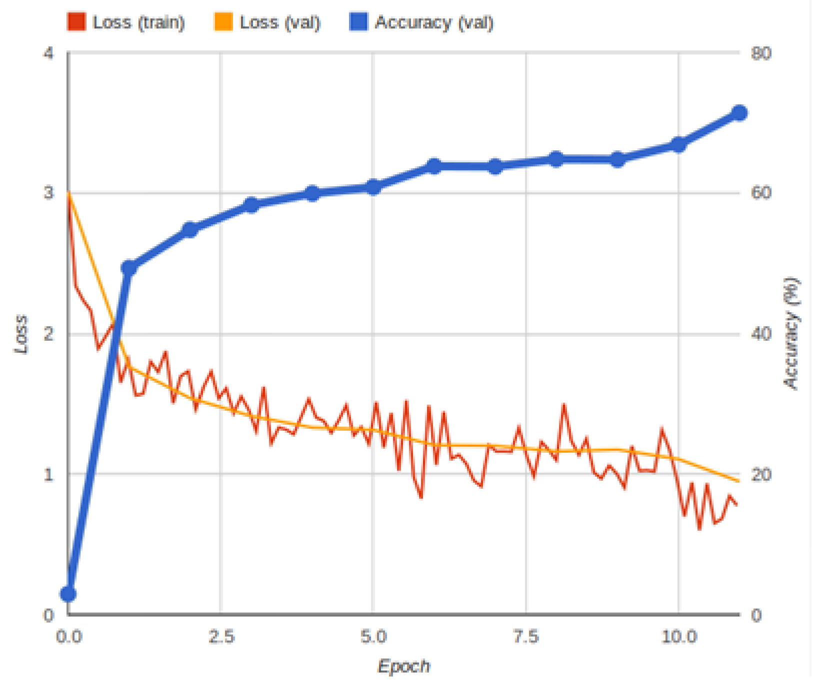


Нейросеть



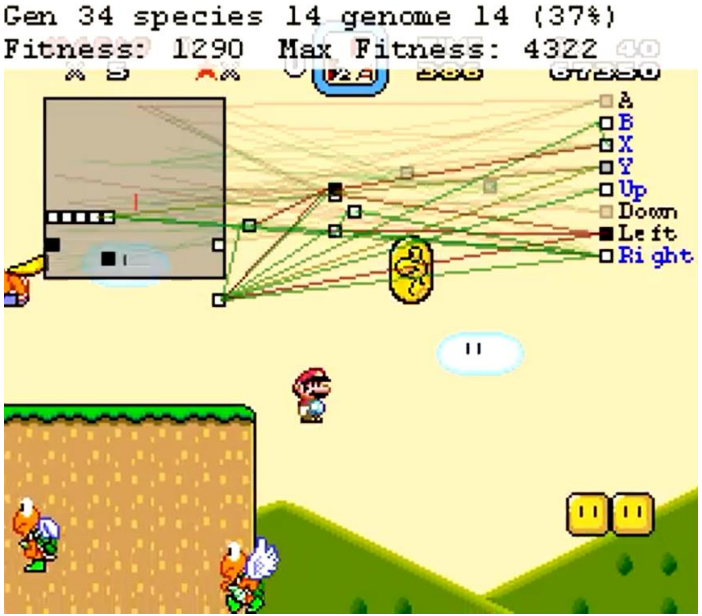
Обучение нейросети

- Stochastic
  - После каждого примера следует стадия обновления весов
  - Добавляет шум в метод градиентного спуска. Использует локальный градиент только одного пример Снижает шанс остановки в локальном минимуме
- Batch
  - Множество примеров обрабатываются по очереди, в результате чего происходит накопление ошибки для всего набора данных
  - Обычно приводит к большей стабильности и лучшей сходимости
- Mini-batch
  - Разумный компромисс



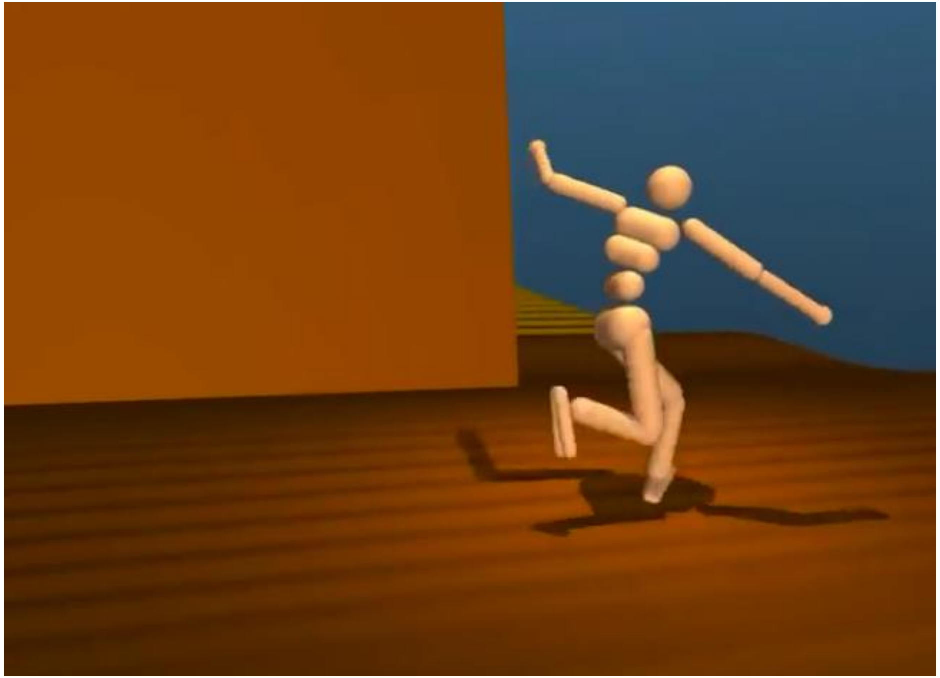
Виды обучения

- Обучение с учителем
  - Нейросеть получает примеры соответствия входных данных выходным
  - Между ними может быть некоторая зависимость, которая неизвестна
- Обучение без учителя
  - Система спонтанно обучается выполнять поставленную задачу без вмешательства со стороны экспериментатора.
  - Как правило, это пригодно только для задач, в которых известны описания множества объектов
- Обучение с подкреплением
  - Частный случай обучения с учителем, где реакция среды служит откликом на действия сети



Google DeepMind

- https://www.youtube.com/watch?v=gn4nRCC9TwQ



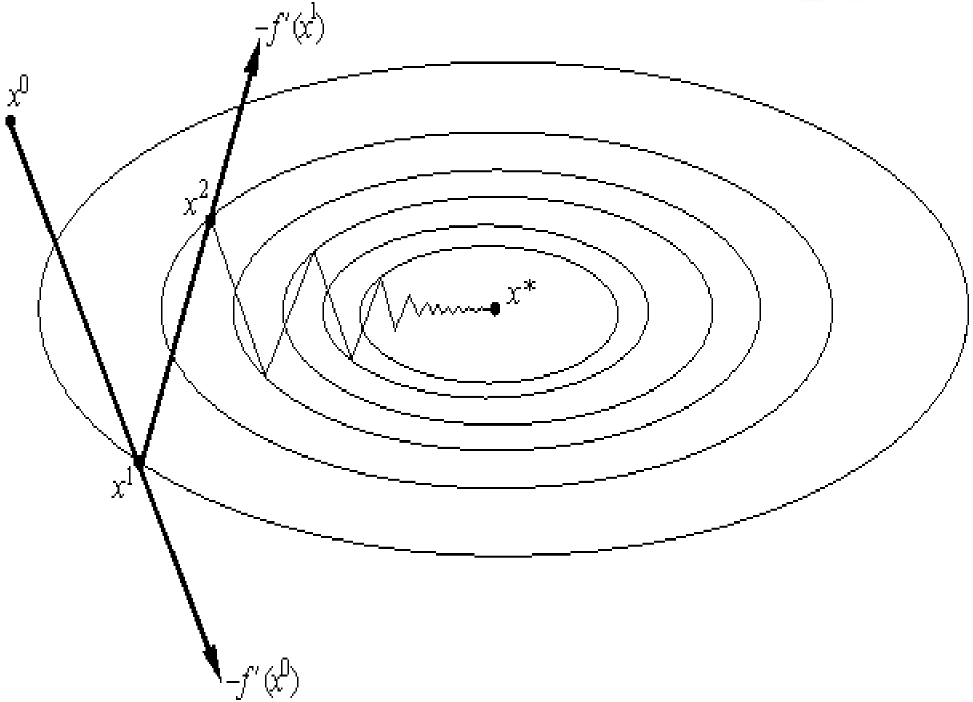
GD

- Производные:

$$\frac{d(e^u)}{dx} = e^u \frac{du}{dx}$$

$$\frac{d(g+h)}{dx} = \frac{dg}{dx} + \frac{dh}{dx}$$

$$\frac{d(g^n)}{dx} = ng^{n-1} \frac{dg}{dx}$$

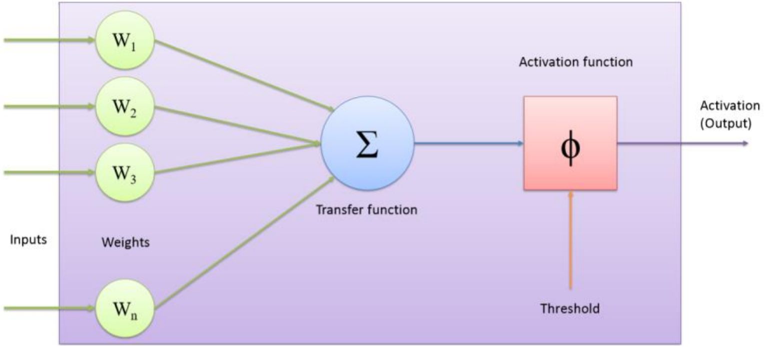


Backpropagation

$$E = \frac{1}{2}\sum_k (t_k - a_k)^2$$

$$\Delta w_{kj} \propto -\frac{\partial E}{\partial w_{kj}}$$

$$\Delta w_{kj} = -\varepsilon \frac{\partial E}{\partial a_k} \frac{\partial a_k}{\partial net_k} \frac{\partial net_k}{\partial w_{kj}}$$



Backpropagation

$$\frac{\partial E}{\partial a_k} = \frac{\partial(\frac{1}{2}(t_k - a_k)^2)}{\partial a_k} = -(t_k - a_k)$$

$$\frac{\partial a_k}{\partial net_k} = \frac{\partial(1 + e^{-net_k})^{-1}}{\partial net_k} = \frac{e^{-net_k}}{(1 + e^{-net_k})^2}$$

$$1 - \frac{1}{1 + e^{-net_k}} = \frac{e^{-net_k}}{1 + e^{-net_k}}$$

$$a_k(1 - a_k)$$

Backpropagation

$$\frac{\partial net_k}{\partial w_{kj}} = \frac{\partial (w_{kj} a_j)}{\partial w_{kj}} = a_j$$

$$\Delta w_{kj} = \varepsilon \overbrace{(t_k - a_k) a_k (1 - a_k)}^{\delta_k} a_j$$

Backpropagation

$$\Delta w_{ji} \propto -[\sum_k \frac{\partial E}{\partial a_k} \frac{\partial a_k}{\partial net_k} \frac{\partial net_k}{\partial a_j}] \frac{\partial a_j}{\partial net_j} \frac{\partial net_j}{\partial w_{ji}}$$

$$= \varepsilon[\sum_k \overbrace{(t_k - a_k) a_k (1 - a_k)}^{\delta_k} w_{kj}] a_j (1 - a_j) a_i$$

$$= \varepsilon \overbrace{[\sum_k \delta_k w_{kj}] a_j (1 - a_j)}^{\delta_j} a_i$$

$$\Delta w_{ji} = \varepsilon \delta_j a_i$$

Производные