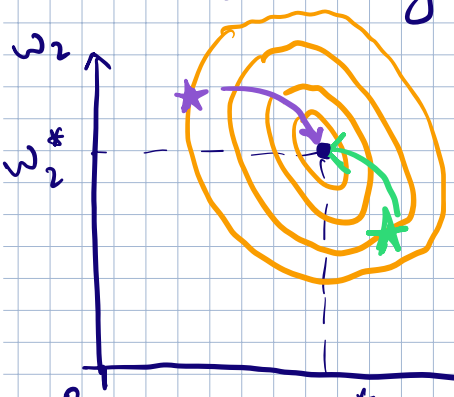
Loss landscape

$$f(w) = \frac{1}{e}\|Xw - y\|_2^2 + \frac{\lambda}{2}\|w\|_2^2 \to \min_w$$ - lin regression



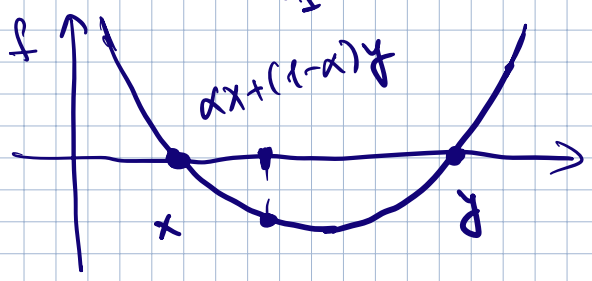
Logistic regression:

$$\frac{1}{e}\sum_{i=1} \log\left(1 + e^{-y_i \langle w, x_i \rangle}\right) + \frac{\lambda}{2}\|w\|_2^2$$

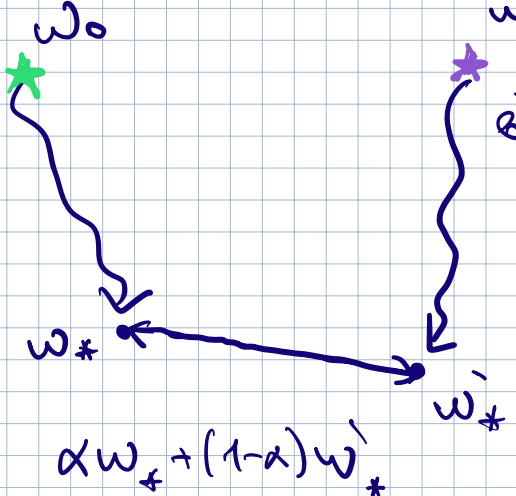
Convex function f

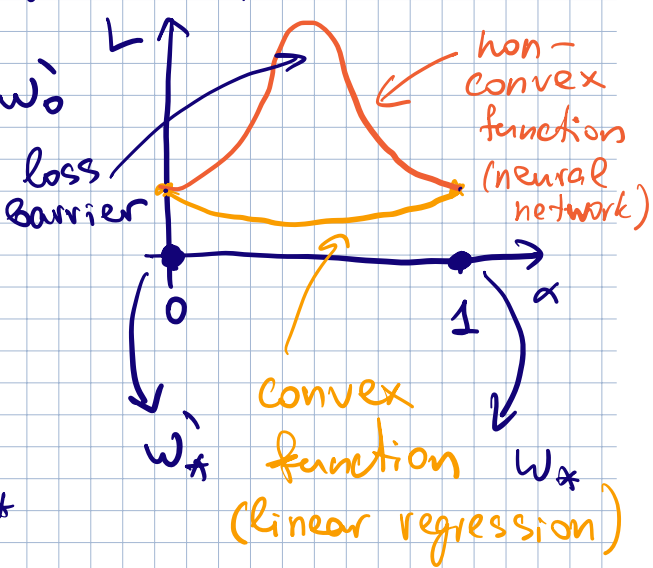
$$f(\alpha x + (1-\alpha)y) \le \alpha f(x) + (1-\alpha) f(y)$$

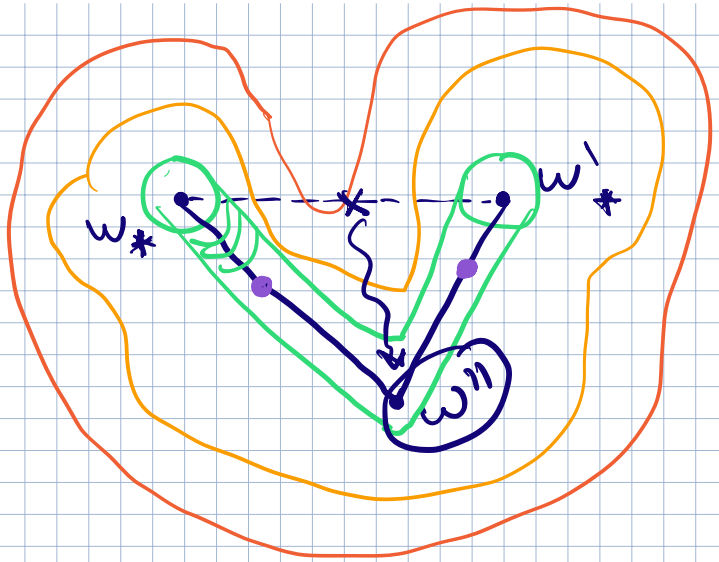
$0 \le \alpha \le 1$



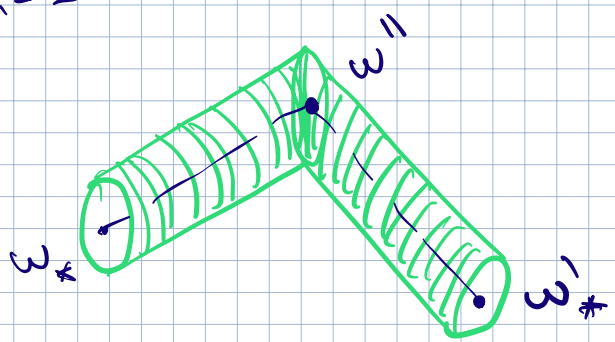
Neural networks are not convex $w \in \mathbb{R}^d$

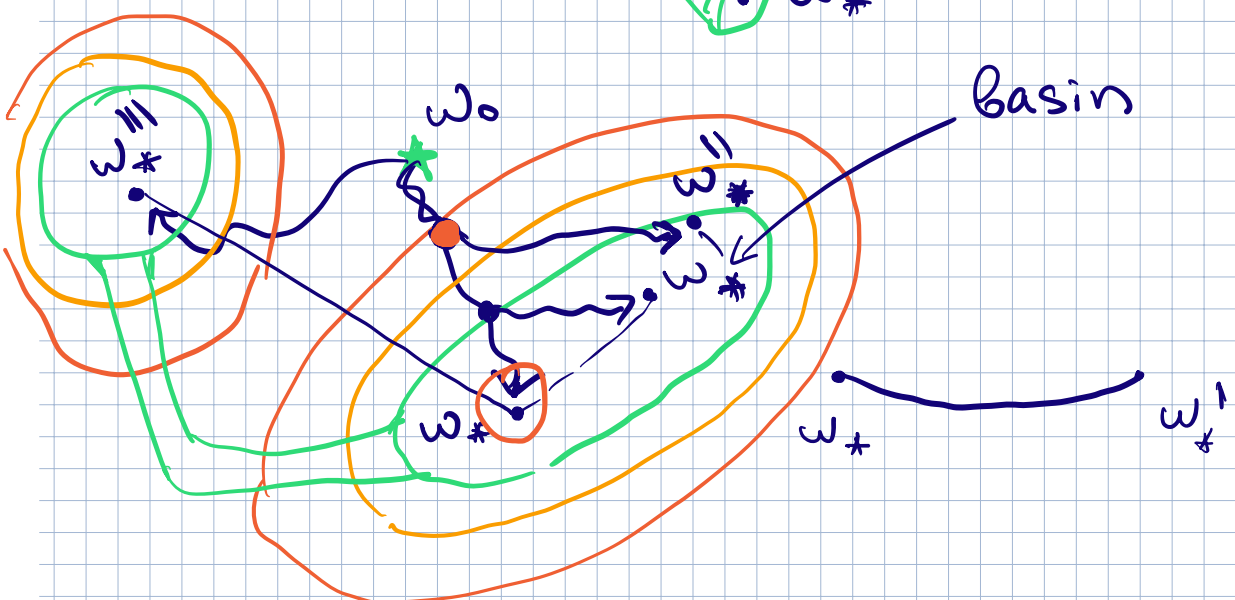




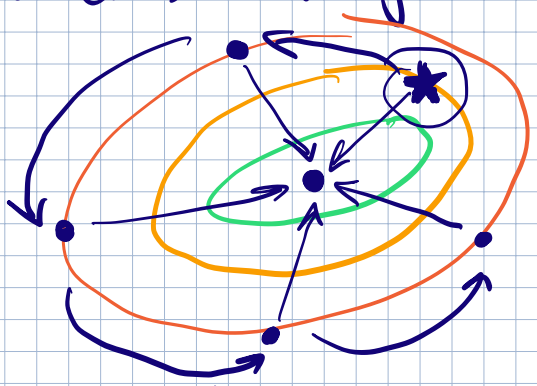


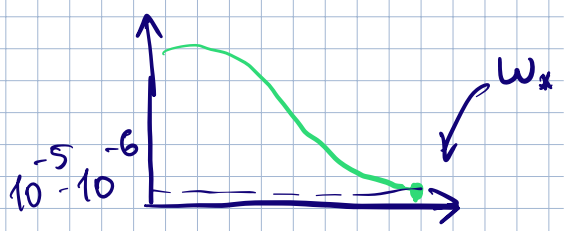
$$\mathbb{E}_{\substack{\alpha,\beta \\ \in U[0,1]}} \left[L(\alpha w_* + (1-\alpha) w'') + L(\beta w'' + (1-\beta) w'_*) \right] \to \min_{w''}$$



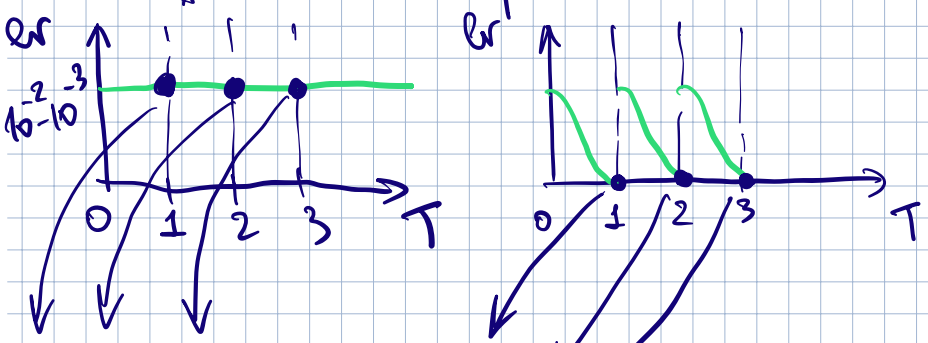


Stochastic weight averaging (SWA)



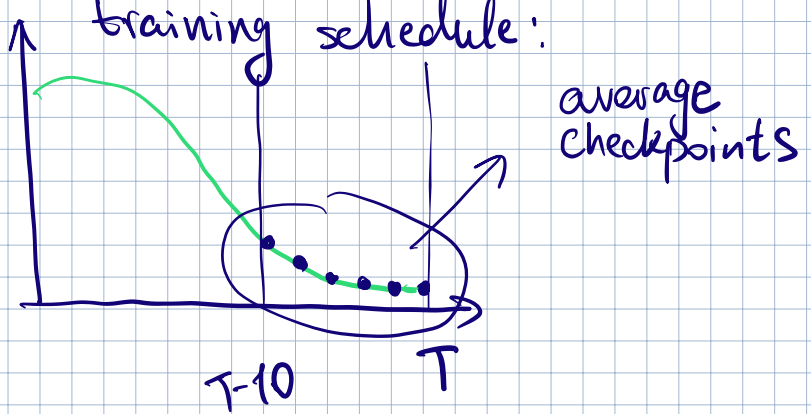


1. w_* – trained optimum



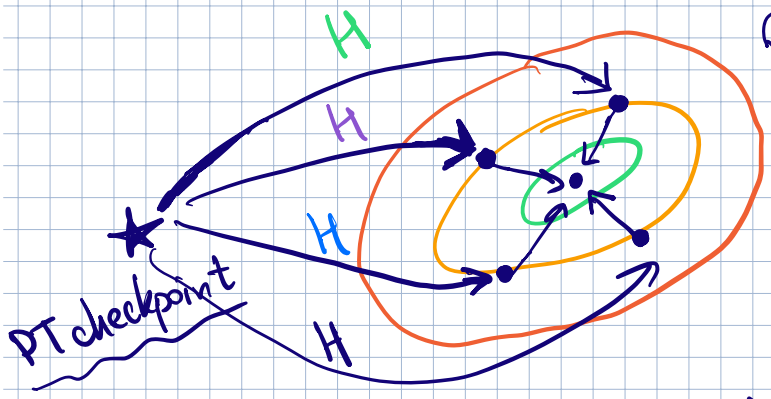
Save checkpoints

2. training schedule:



- Update Batch Norm stats
  forward pass epoch on training data

Pre-trained checkpoint → fine-tune



$\{w_1, \ldots, w_K\}$

$\{i_1, \ldots, i_n\} \subseteq \{1, \ldots, K\}$

validation set

$$w := \frac{1}{n} \sum_{j=1}^{n} w_{i_j}$$

Greedy search

Greedy model soup > Uniform model soup