#nodes in hidden layers

#hidden layers

activation fns

Examples of "hyperparameters"

not trainable parameters

- NN basically repeated matrix multiplication
  + vectorized activations
  \[ \rightarrow \text{perfect for GPU acceleration!} \]

- Counting parameters

  let's say \( d = 100 \), \( L = 10 \), \( n = 1000 \)

  \[ \dim W^{(l)} \sim (10^3)^2 \sim 10^6 \]

  \( a \times 10 \) hidden layers

  \( \sim 10^7 \) parameters

  \[ \rightarrow \# \text{ of NN parameters can easily get very large!} \]

  can still be fit to data using powerful algorithms & hardware

  \[ \rightarrow 1^{st} \text{ demo n.b.: contracting simple MLP} \]
  in Keras & in pytorch
Training NNs

How to fit \( f(x; \theta) \) with many (e.g. 1M) parameters?

Want to minimize \( L(\theta) \) wrt \( \theta \)

\[
\frac{\partial L}{\partial \theta} = 0 \quad (\text{ideally) global minimum}
\]

\( L \) might have many local minima or saddles

\[
\downarrow \quad \text{don't want to get stuck}
\]

Many approaches...

- One idea: Newton method

\[
\theta = \theta_0 \quad \text{initial guess}
\]

\[
L(\theta) = L(\theta_0) + \frac{1}{2} \delta \theta^T \nabla^2 L \delta \theta + \nabla L \delta \theta
\]

- Minimize 2nd order approx \( \rightarrow 8 \theta \)

- Repeat for \( \theta + 8 \theta \)

- Hopefully converge to \( \hat{\theta} \)

- Why is this infeasible!? \n
\[
\rightarrow \frac{\partial^2 L}{\partial \theta^2} \quad \text{too high don't need to invert it!}
\]
Instead, NNs are optimized via first order methods:

\[ \delta \theta = \eta \frac{\partial L}{\partial \theta} \]

"gradient descent"

Idea: more induction where \( L \) decreases

\( \eta \) = "learning rate" — important hyperparameter, often set to fine-tune for best performance.

Problem: GD can get stuck in local minima

\[ \frac{\partial L}{\partial \theta} = 0 \Rightarrow \delta \theta = 0 \]

no update

get stuck at \( \theta = \theta_0 \)

\[ \rightarrow \text{"stochastic gradient descent"} \]

Noisy gradients are better than true gradients!

How to get noisy gradients?