Neural networks

Deep learning - difficulty of training
Topics: multilayer neural network

- Could have \( L \) hidden layers:
  - layer input activation for \( k > 0 \) (\( h^{(0)}(x) = x \))
    \[ a^{(k)}(x) = b^{(k)} + W^{(k)} h^{(k-1)}(x) \]
  - hidden layer activation (\( k \) from 1 to \( L \)):
    \[ h^{(k)}(x) = g(a^{(k)}(x)) \]
  - output layer activation (\( k=L+1 \)):
    \[ h^{(L+1)}(x) = o(a^{(L+1)}(x)) = f(x) \]
Topics: why training is hard

- First hypothesis: optimization is harder (underfitting)
  - vanishing gradient problem
  - saturated units block gradient propagation

- This is a well known problem in recurrent neural networks
**Topics:** why training is hard

- Second hypothesis: overfitting
  - we are exploring a space of complex functions
  - deep nets usually have lots of parameters
- Might be in a high variance / low bias situation
Topics: why training is hard

• Second hypothesis: overfitting
  ‣ we are exploring a space of complex functions
  ‣ deep nets usually have lots of parameters

• Might be in a high variance / low bias situation
DEEP LEARNING

Topics: why training is hard

- Depending on the problem, one or the other situation will tend to dominate

- If first hypothesis (underfitting): use better optimization
  - this is an active area of research

- If second hypothesis (overfitting): use better regularization
  - unsupervised learning
  - stochastic «dropout» training