Neural Networks

Deep Neural Network for 3-way Classification

$$h_i^{\text{layer } l} = \phi\left(\sum_j w_{ji}^{(\text{layer } l)} \cdot h_j^{(\text{layer } l-1)}\right)$$

- L := hidden layers - $h^{(l)}$:= activations at layer l - $w^{(l)}$:= weights taking activations from layer $l-1$ to l - ϕ := nonlinear activation function - Must be nonlinear because otherwise hidden layers are collapsed - Last layer is still just a logistic regression - Prev layers just learn the 'features' of input

Rectified Linear Unit (ReLU)

$$\phi(z) = \max(0, z)$$

- Type of nonlinear activation function - Commonly used

Training Neural Networks

- Just like logistic regression:

$$\max_w ll(w) = \max_w \sum_i \log \mathbb{P}(y^{(i)}|x^{(i)}; w)$$

- Just w tends to be a much larger vector - Just run gradient ascent and stop when log likelihood of hold-out data starts to decrease - Algorithm: 1. Initialize w 2. Repeat: $w \leftarrow w + \alpha * \sum_i \nabla \log \mathbb{P}(y^{(i)}|x^{(i)}; w)$ - α := learning rate (generally small)

Computing Derivatives

- Automatic differentiation exists
- Relatively quick with backpropagation

Neural Network Properties

- Theorem (Universal Function Approximators) := a two-layer neural network with a sufficient number of neurons can approximate any continuous function to any desired accuracy
- Can be seen as learning the features
- Large number of neurons can cause overfitting

Preventing Overfitting

- Early stopping ≜ stop training when accuracy on held out data set starts going down
- Weight Regularization ≜ add an objective term to penalize weight magnitude
  - Good because w can grow without constraint
  - Use a constraint hyperparameter λ (typically 0.1 to 0.0001 or smaller)

$$\max_w \sum_i \log \mathbb{P}(y^{(i)}|x^{(i)}; w) - \frac{\lambda}{2} \sum_j w_j^2$$

    ## Simplicity
- Reduce the hypothesis/model space
  - Assume more
  - Fewer features or neurons
  - Other limits on model structure
- Regularization

- Laplace smoothing
- Weight regularization
- Hypothesis state stays big, but harder to get to outskirts