CSE 252A: Deep Networks II

Lecturer: David Kriegman

Scribed by Owen Jow on December 04, 2018

1 Neural Networks for Classification

Typically, we'll have a neural network output a vector containing the posterior probabilities $P(\omega_i \mid \mathbf{x})$ of the input $\mathbf{x}$ belonging to each class ω_i. This is the same as what the Bayes classifier outputs.

2 Regularization

We use regularization to avoid overfitting (and hopefully improve a network's ability to generalize), typically making our model smoother and/or reducing the number of ways it can solve the problem. An example is dropout.

3 Convolutional Networks

In CNNs, we usually prefer max pooling to average pooling for dimensionality reduction[1] because max pooling is nonlinear while average pooling can already be done by the convolutions.

A 2D convolution uses a 4D (`height`) $\times$ (`width`) $\times$ (`in channels`) $\times$ (`out channels`) kernel because each application of the kernel takes a (`height`) $\times$ (`width`) $\times$ (`in channels`) volume and produces an (`out channels`) vector. This happens at each 2D location in the input. Another way of looking at it, of course, is that we convolve the input with (`out channels`) different 3D (`height`) $\times$ (`width`) $\times$ (`in channels`) kernels and end up with that many feature maps.

[1] good because reduces number of parameters, combats overfitting, limits sensitivity to exact locations of features