

Deep Learning - Foundations and Concepts

Chapter 6. Deep Neural Networks

nonlineark@github

February 18, 2025

Outline

1 Limitations of Fixed Basis Functions

The curse of dimensionality

In spaces of higher dimensionality, the number of combinations of values must be considered could be huge. This effect is known as combinatorial explosion:

- A polynomial regression of order M for a single input variable needs $M + 1$ parameters. If there are D input variables, the number of parameters needed will be D^{M+1} .
- The histogram based classification for 1-dimensional input needs N buckets. If the input is D -dimensional, the number of buckets needed will be N^D .

For a machine learning model, this usually means that the amount of data needed to generalize accurately grows exponentially.

High-dimensional spaces

High-dimensional spaces can defeat one's geometrical intuitions:

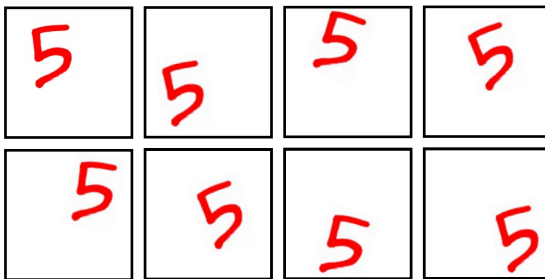
- In spaces of high dimensionality, most of the volume of a hypersphere is concentrated in a thin shell near the surface.
- In spaces of high dimensionality, the probability mass of the Gaussian is concentrated in a thin shell at a specific radius (a soap bubble).

Data manifolds

Although data may be in high-dimensional spaces, real data will generally be confined to a region of the data space having lower effective dimensionality. Effectively, neural networks learn a set of basis functions that are adapted to data manifolds.

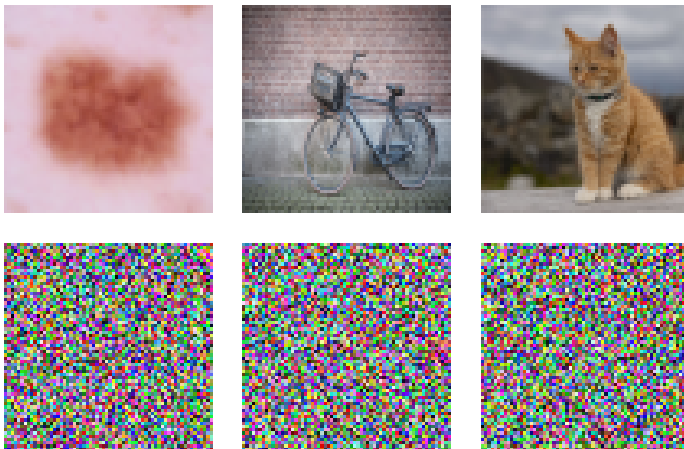
Data manifolds

Figure: Images of a handwritten digit that lives on a nonlinear three-dimensional manifold



Data manifolds

Figure: Natural images vs. randomly generated images



Data-dependent basis functions

- Simple basis functions that are chosen independently of the problem being solved can run into significant limitations.
- Using expert knowledge to hand-craft the basis functions was superseded by data-driven approaches in which basis functions are learned from the training data.
- Methods such as radial basis functions and support vector machines have been superseded by deep neural networks, which are much better at exploiting very large data sets efficiently.