# Deep Learning

## Chapter 9: Regularized Autoencoders

**Mina Rezaei**

Department of Statistics – LMU Munich

Winter Semester 2020

# MANIFOLD LEARNING

- **Manifold hypothesis**: Data of interest lies on an embedded non-linear manifold within the higher-dimensional space.
- A **manifold**:
  - is a topological space that locally resembles the Euclidean space.
  - in ML, more loosely refers to a connected set of points that can be approximated well by considering only a small number of dimensions.

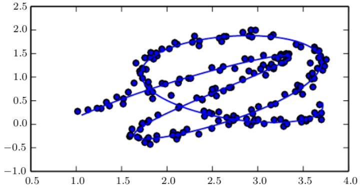


**Figure:** from Goodfellow et. al

# MANIFOLD LEARNING

- An important characterization of a manifold is the set of its tangent planes.
- **Definition**: At a point **x** on a *d*-dimensional manifold, the **tangent plane** is given by *d* basis vectors that span the local directions of variation allowed on the manifold.

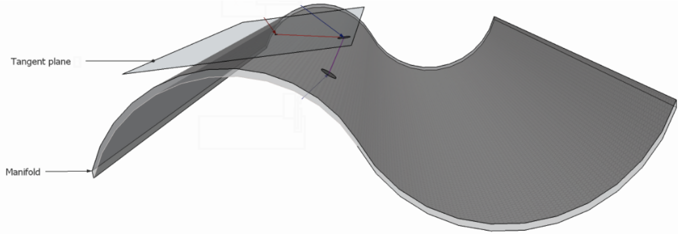


**Figure:** A pictorial representation of the tangent space of a single point, $\boldsymbol{x}$, on a manifold (Goodfellow et al. (2016)).

# MANIFOLD LEARNING

- Manifold hypothesis does not need to hold true.
- In the context of AI tasks (e.g. processing images, sound, or text) it seems to be at least approximately correct, since :
    - probability distributions over images, text strings, and sounds that occur in real life are highly concentrated (randomly sampled pixel values do not look like images, randomly sampling letters is unlikely to result in a meaningful sentence).
    - samples are connected to each other by other samples, with each sample surrounded by other highly similar samples that can be reached by applying transformations (E.g. for images: Dim or brighten the lights, move or rotate objects, change the colors of objects, etc).

# LEARNING MANIFOLDS WITH AES

- AEs training procedures involve a compromise between two forces:
  1. Learning a representation $\boldsymbol{z}$ of a training example $\mathbf{x}$ such that $\mathbf{x}$ can be approximately recovered from $\boldsymbol{z}$ through a decoder.
  2. Satisfying an architectural constraint or regularization penalty.
- Together, they force the hidden units to capture information about the structure of the data generating distribution
- important principle: AEs can afford to represent only the variations that are needed to reconstruct training examples.
- If the data-generating distribution concentrates near a low-dimensional manifold, this yields representations that implicitly capture a local coordinate system for the manifold.

# LEARNING MANIFOLDS WITH AES

- Only the variations tangent to the manifold around $\mathbf{x}$ need to correspond to changes in $\boldsymbol{z} = enc(\mathbf{x})$. Hence the encoder learns a mapping from the input space to a representation space that is only sensitive to changes along the manifold directions, but that is insensitive to changes orthogonal to the manifold.

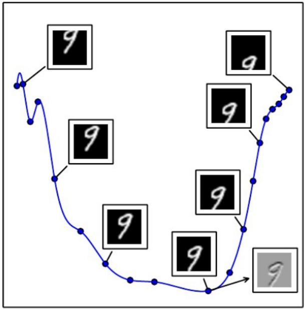

**Figure:** from Goodfellow et al. (2016)

# LEARNING MANIFOLDS WITH AES

- Common setting: a representation (embedding) for the points on the manifold is learned.
- Two different approaches
  1. Non-parametric methods: learn an embedding for each training example.
  2. Learning a more general mapping for *any* point in the input space.
- AI problems can have very complicated structures that can be difficult to capture from only local interpolation.
  $\Rightarrow$ Motivates use of **distributed representations** and deep learning!