# Tangent Distance, Tangent Prop and Manifold Tangent Classifier

Sargur N. Srihari
srihari@cedar.buffalo.edu

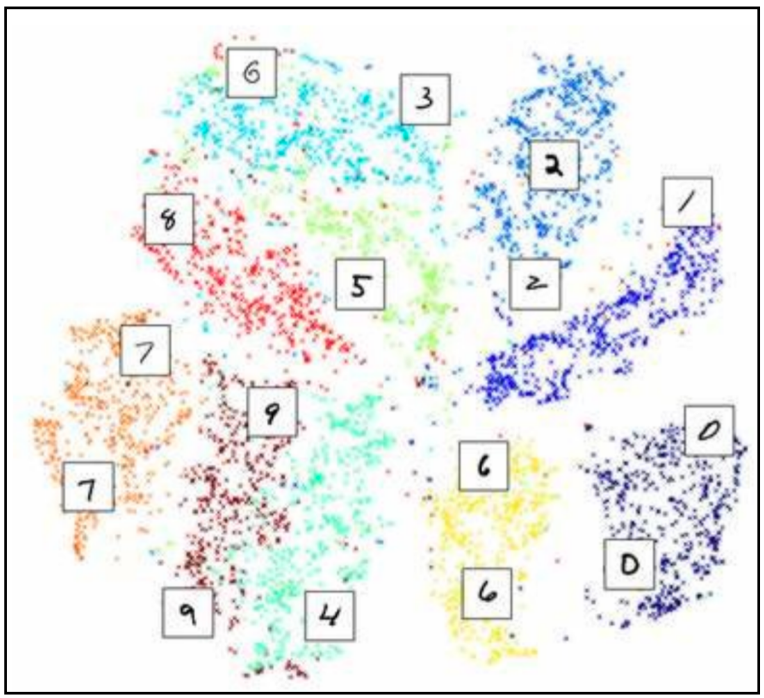
# Topics

1. Manifolds in data space
2. Tangent distance
3. Tangent propagation regularization
4. Manifold Tangent Classifier
5. Manifold learning

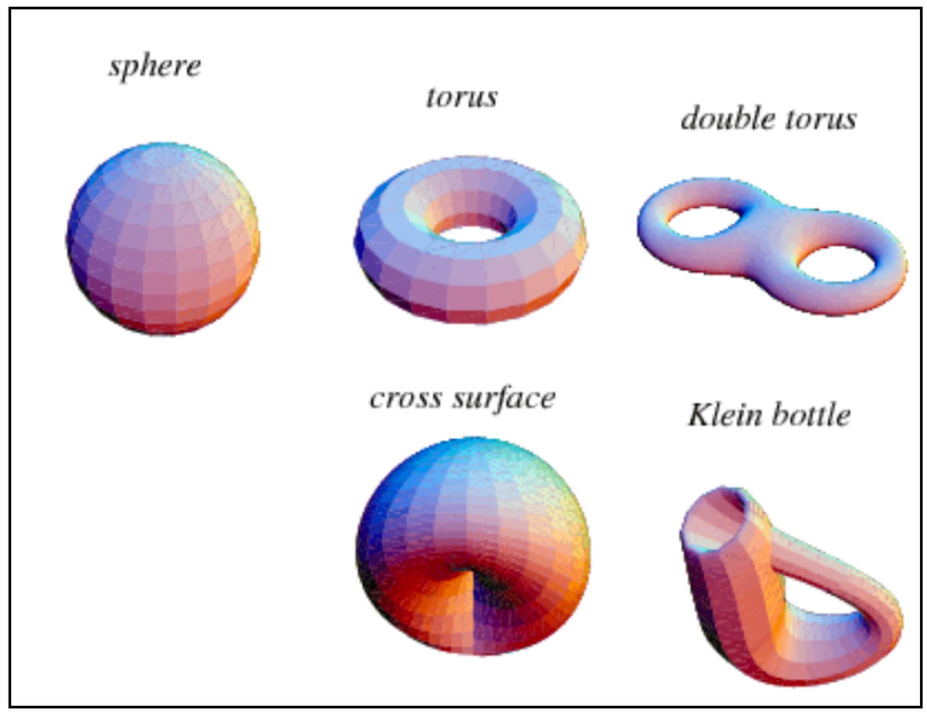
# 1. Manifolds in Data Space

- Data lies on low dimensional manifolds

Manifolds of handwritten digits in 2-D




Common examples of manifolds

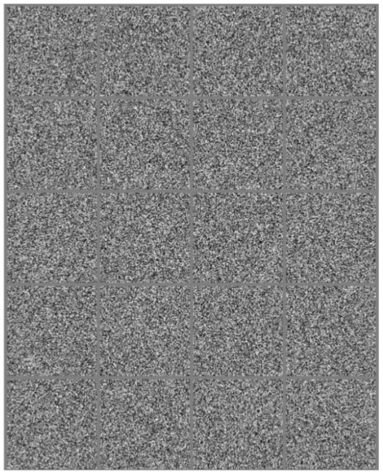



Note that the manifold is only the surface of these objects and not the interior

# Manifold hypothesis justification

- Images
  - Distributions are highly concentrated
  - Uniformly sampled points

    look like static noise, never structured



    - Although non-zero probability of generating a face, it is never observed
- Text
  - If you generate a document by randomly generating text, near zero probability of generating meaningful text
  - Natural language sequences occupy a small volume of total space of sequences of letters

# Data manifolds are smooth

- Common theme of these examples:
  - They are somewhat smooth
    - Meaning that there are no sharp spikes or edges
- Overall shape of manifold can be amorphous
  - Describe datasets that don’t have rigid boundaries

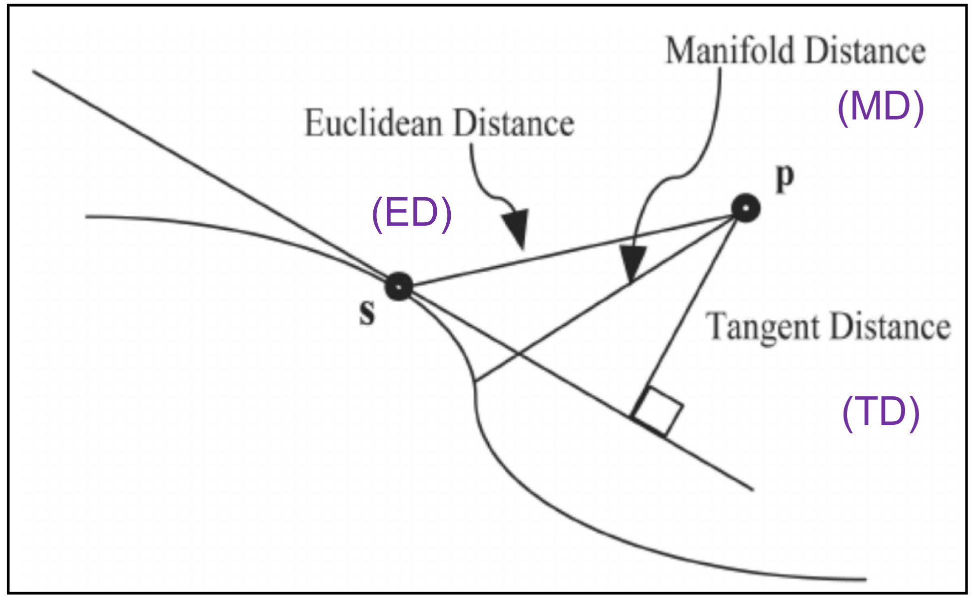
# Manifolds traced by transformations

- Manifolds can be traced by making small transformations
- Manifold structure of a dataset of human faces

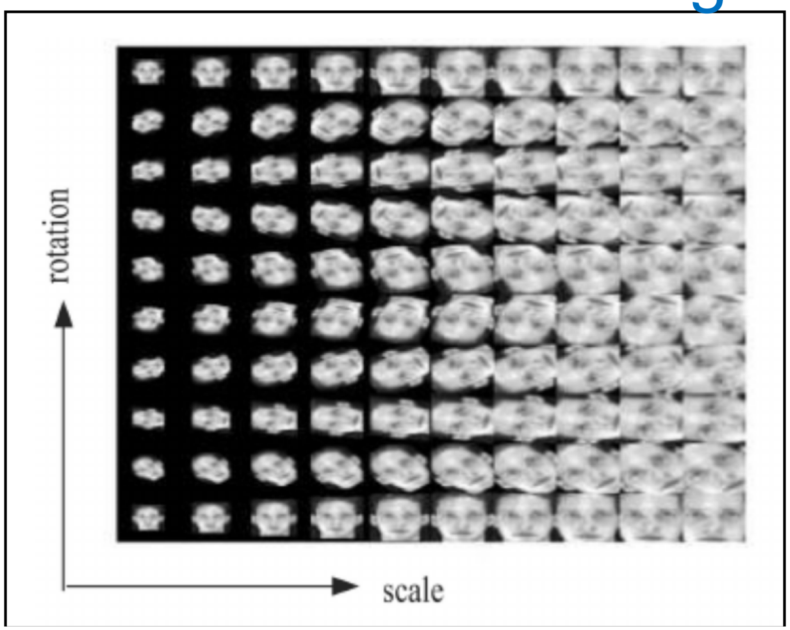
# Tangent Distance

Three ways to measure distance from a point to a manifold:

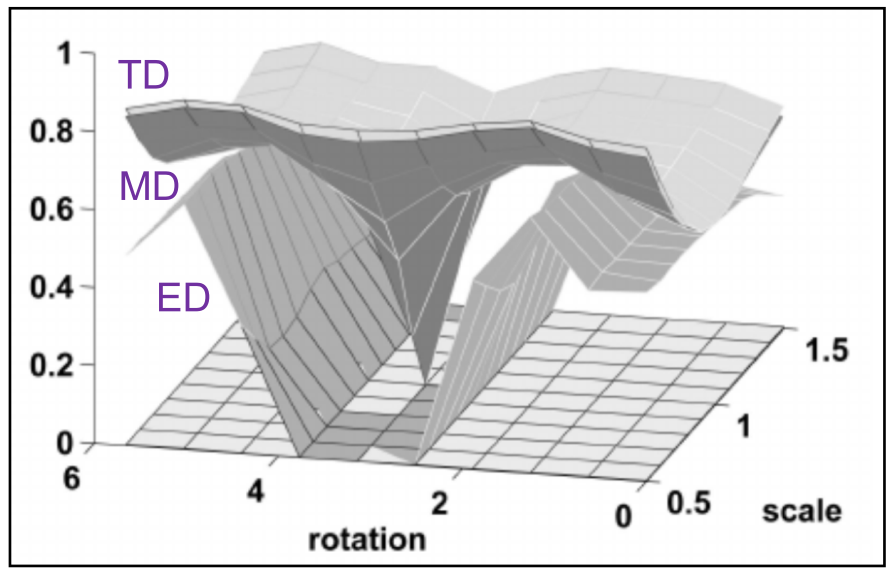
Manifold distance (MD)
Euclidean distance (ED)
Tangent distance (TD)




## Transformed face images




## Distance between s and p



# Tangent vector from finite differences

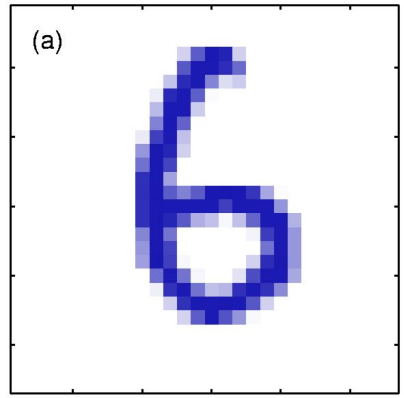
- Tangent vector $\tau_n$

$$\tau_n = \left.\frac{\partial s(x_n,\xi)}{\partial \xi}\right|_{\xi=0}$$

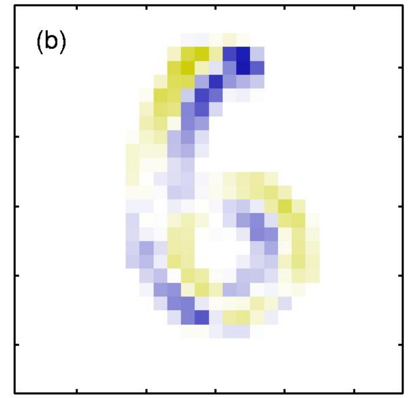
- Approximation using finite differences

  By subtracting original vector $x_n$ from the corresponding vector after transformations using a small value of ξ and then dividing by ξ

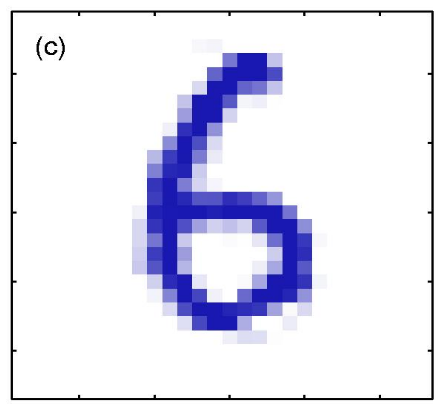
Original Image $x_n$



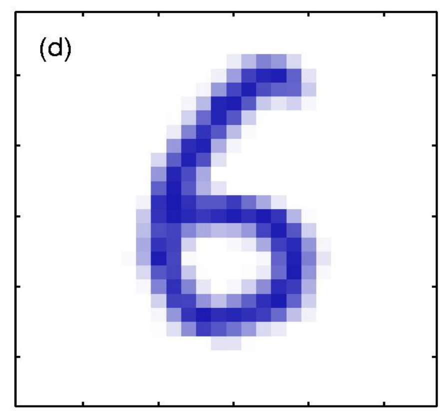

Tangent vector $\tau_n$ corresponding to small clockwise rotation




Adding small contribution from tangent vector to image $x_n+\varepsilon\tau_n$




True image rotated for comparison



# Tangent distance classifier

- Tangent Distance is used to build invariance properties into distance-based methods such as *nearest-neighbor* classifiers
- Distance between point $\boldsymbol{x}$ and manifold is computationally difficult
  - Would require solving an optimization problem
- Cheap alternative is to use distance between point and tangent plane
  - By solving a linear system in the dimension of the manifold
    - Requires specifying the tangent vector

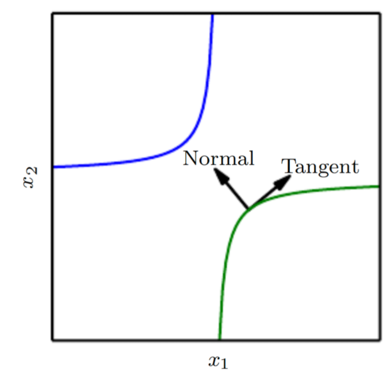
# 3. Tangent Propagation

- Related to Tangent Distance
- Trains a classifier with an extra penalty
- Makes each output $f(\boldsymbol{x})$ of the neural network locally invariant to known factors of variation
- These factors of variation correspond to movement along the manifold near which examples of the same class concentrate

# Tangent Prop Regularization

- Local invariance is achieved by requiring $\nabla_{\boldsymbol{x}} f(\boldsymbol{x})$ to be orthogonal to the known manifold tangent vectors $\boldsymbol{v}^{(i)}$ at $\boldsymbol{x}$
- Equivalently that the directional derivative of $f$ at $\boldsymbol{x}$ in the direction of $\boldsymbol{v}^{(i)}$ be small by adding a regularization penalty $\Omega$

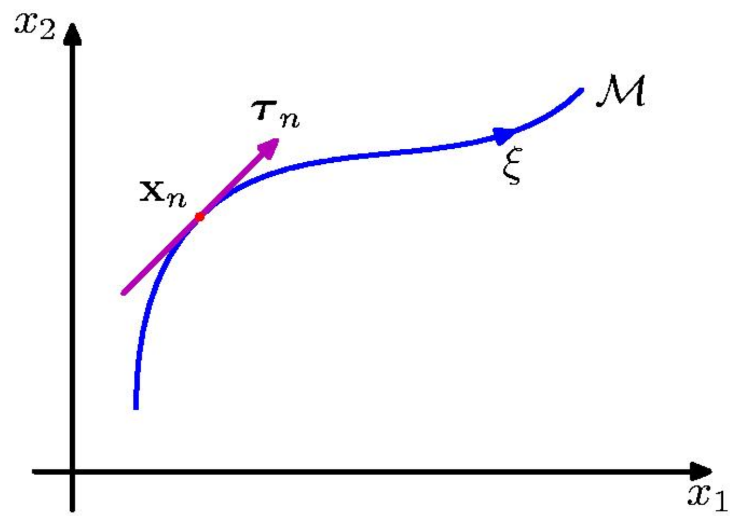
$$\Omega(f) = \sum_i \left( \left( \nabla_{\boldsymbol{x}} f(\boldsymbol{x}) \right)^T \boldsymbol{v}^{(i)} \right)^2$$




- The regularizer can be scaled using a hyperparameter $\lambda$
  – Also need to sum over many outputs

# Tangent Prop on 1-D manifold

- Regularization to encourage invariance to transformations
  - by technique of tangent propagation
- Consider effect of transformation on input $\boldsymbol{x}_n$
  - A $1$-$D$ continuous transformation parameterized by ξ applied to $\boldsymbol{x}_n$ sweeps a manifold $\mathcal{M}$ in $D$-dimensional input space




Two-dimensional input space showing effect of continuous transformation with single parameter ξ
Let the vector resulting from acting on $\boldsymbol{x}_n$ by this transformation be denoted by $s(\boldsymbol{x}_n,\xi)$ defined so that $s(\boldsymbol{x},0)=\boldsymbol{x}$.
Then the tangent to the curve $\mathcal{M}$ is given by the directional derivative $\tau = \delta s/\delta\xi$ and the tangent vector at point $\boldsymbol{x}_n$ is given by

$$\boldsymbol{\tau}_n = \left.\frac{\partial s(x_n,\xi)}{\partial \xi}\right|_{\xi=0}$$

# Tangent Prop Regularization

- Under a transformation of input vector
  - The network output vector will change
  - Derivative of output $k$ wrt $\xi$ is given by

$$\left.\frac{\partial y_k}{\partial \xi}\right|_{\xi=0} = \sum_{i=1}^{D} \frac{\partial y_k}{\partial x_i} \left.\frac{\partial x_i}{\partial \xi}\right|_{\xi=0} = \sum_{i=1}^{D} J_{ki} \tau_i$$

    - where $J_{ki}$ is the $(k,i)$ element of Jacobian Matrix $J$
- Result used to modify standard error function

$$\tilde{E} = E + \lambda \Omega$$

where $\lambda$ is a regularization coefficient and

$$\Omega = \frac{1}{2} \sum_n \sum_k \left( \left.\frac{\partial y_{nk}}{\partial \xi}\right|_{\xi=0} \right)^2 = \frac{1}{2} \sum_n \sum_k \left( \sum_{i=1}^{D} J_{nki} \tau_{ni} \right)^2$$

    - To encourage local invariance in neighborhood of data pt

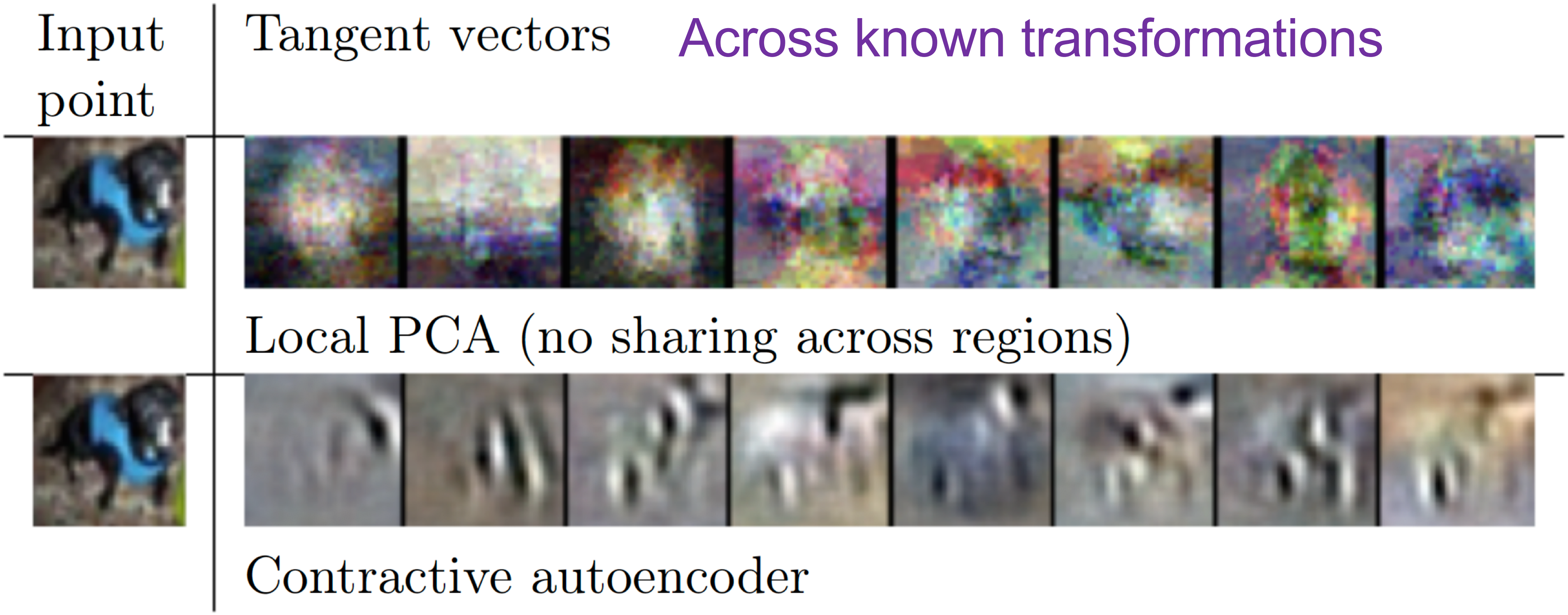
# Tangent Prop Implementation

- As with tangent distance tangent vectors are derived apriori
- From the formal knowledge of the effect of transformations such as translation, rotation and scaling of images
- Tangent prop has been used for:
  - Supervised learning
  - Reinforcement learning

# 4. Manifold Tangent Classifier

- Eliminates need to know tangent vectors a priori
- Uses autoencoders to estimate manifold tangent vectors
- Goes beyond classical invariants that arise from image geometry (translate, rotate, scale)
- Include factors that are object specific such as moving body parts
- Example next

# Autoencoder specified tangent vectors



# 5. Manifold Learning

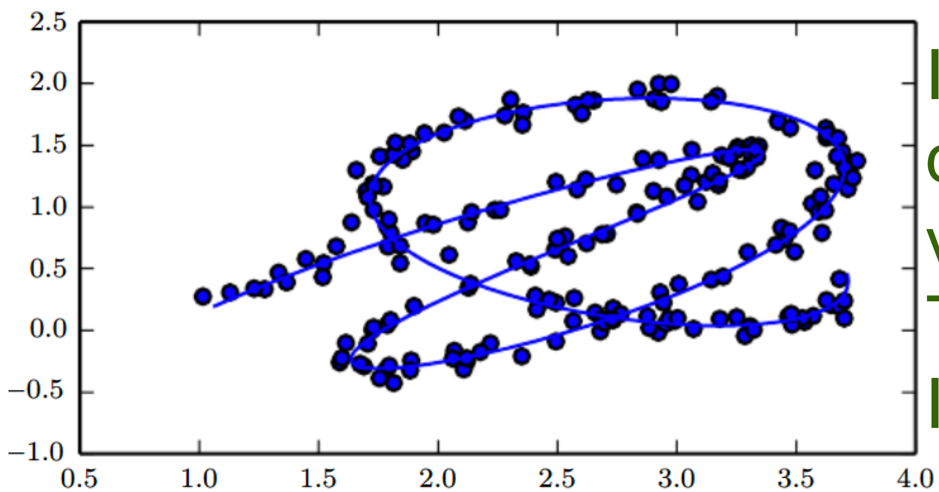
- An important idea underlying many ideas in machine learning
- A manifold is a connected region
  - Mathematically it is a set of points in a neighborhood
  - It appears to be in a Euclidean space
    - E.g., we experience the world as a 2-D plane while it is a spherical manifold in 3-D space

# Manifold in Machine Learning

- Although manifold is mathematically defined, in machine learning it is loosely defined:
  - A connected set of points that can be approximated well by considering only a small no of degrees of freedom embedded in a higher-dimensional space

Training data lying near a 1-D Manifold in a 2-D space

The solid line indicates the underlying manifold that the learner should infer




In machine learning we allow the dimensionality of the manifold to vary from one point to another. This often happens when a manifold Intersects itself, as in a figure-eight

# Manifold learning surmounts $\mathrm{R}^n$

- It is sometimes hopeless to learn functions with variations across all of $\mathrm{R}^n$
- Manifold learning algorithms surmount this obstacle by assuming most of $\mathrm{R}^n$ consists of invalid inputs
  - And that intersecting inputs occur only along the manifolds
- Introduced for continuous data and in unsupervised learning, the probability concentration idea can be generalized to discrete and unsupervised settings

# Manifolds discovered for Human Faces

- Variational autoencoder discovers underlying two-dimensional coordinate system:
  1. Rotation
  2. Emotion