MANIFOLDS

- THE DATA-GENERATING DISTRIBUTION IS ASSUMED TO CONCENTRATE NEAR REGIONS OF LOW DIMENSIONALITY
- **MANIFOLD** IN ML: CONTINUOUS SPACES THAT LOCALLY RESEMBLE EUCLIDEAN SPACES. DATA MAY LIE NEAR THE MANIFOLD, NOT STRICTLY ON IT. DIMENSIONALITY MAY NOT BE THE SAME EVERYWHERE. ALSO DISCRETE SPACES → '**PROBABILITY CONCENTRATION**'. IS CONFIGURATION X PROBABLE?
- PROBABLE CONFIGURATION → GENERALLY SURROUNDED (IN SOME DIRECTIONS) BY OTHER PROBABLE CONFIGS. **LOCAL DIMENSION**: NO. OF INDEPENDENT WAYS TO LOCALLY TRANSFORM CONFIGURATIONS TO OBTAIN OTHER PROBABLES

→ GOOD SOLUTIONS CONCENTRATE ON RIDGES OF HIGH PROBABILITY MASS

A GOOD REPRESENTATION SUCCESSFULLY INVOLVES LEARNING THE PROJECTION OF POINTS ON THE MANIFOLD → **EMBEDDING**, LOWER DIM THAN 'AMBIENT SPACE'. SOME ALGOS DIRECTLY LEARN EMBEDDINGS, OTHERS LEARN ENCODING / PROJECTION FUNCTION

TANGENT PLANES: FOR POINT X ON A D-DIM MANIFOLD, TANGENT PLANE IS D BASIS VECTORS SPANNING THE ALLOWED LOCAL DIRECTIONS OF VARIATION. WE CAN INFINITESIMALLY CHANGE X WHILE STAYING ON THE MANIFOLDS. PANCAKES.

- MANIFOLD LEARNING IS MOSTLY UNSUPERVISED, NONPARAMETRIC METHODS BASED ON NEAREST-NEIGHBOR GRAPHS. TANGENT PLANES ARE D-O-V OF EXAMPLE/NEIGHBOR DIFFERENCE THEN WE FIND GLOBAL COORDINATE SYSTEM THROUGH OPTIMIZATION **ISSUE**: IF MANIFOLD IS NOT VERY SMOOTH → VERY LARGE NO. OF SAMPLES NEEDED TO CAPTURE VARIATIONS BECAUSE LEARNING BY INTERPOLATION
- **BETTER IDEA**: LET'S LEARN AN EXPLICIT OR IMPLICIT COORDINATE SYSTEM FOR THE MANIFOLD. MAIN OBJECTIVE IS TO DISCOVER MANIFOLDS. **EXAMPLE**: VARIATIONAL AUTOENCODERS LEARNING ABOUT ROTATION AND EXPRESSIONS IN FACESPACE

PCA / LINEAR AUTOENCODERS AS MANIFOLD LEARNING

- **ENCODER** $h = f(x) = W^T(x - \mu)$ MINIMIZING RECONSTRUCTION ERROR $E[\|x-\hat{x}\|^2] \to V = W,\ \mu = b = E[x]$ AND ROWS OF W ARE ORTHONORMAL BASIS, EIGENVECTORS OF COV. MATRIX
- **DECODER** $\hat{x} = g(h) = b + Vh$ IN PCA, ROWS ACTUALLY ARE THE EIGENVECTORS, ORDERED → RECONSTRUCTION ERROR IS $\min |R| = \sum_{d+1}^{D} \lambda_i$
- CODE IS COORDINATES OF POINT IN PROJECTED/REDUCED SPACE

SPARSE CODING

- HERE MASS IS CONCENTRATED ON AXIS-ALIGNED SUBSPACES → SETS OF VALUES FOR WHICH MOST AXES ARE 0
- h → BINARY PATTERN β SPECIFYING NONZERO h_i, VARIABLE LENGTH REAL VECTOR $\alpha \in \mathbb{R}^{N_\alpha}$ WITH ACTIVE DIMENSIONS COORDINATES
- β SPECIFIES THE MANIFOLD, GOING THROUGH $x = b$ • MORE RECONSTRUCTION ERROR → MASS BLEEDS OUT, HYPERPLANES ARE PANCAKE-Y
- 2^d HYPERPLANES IN TOTAL BUT MOST INACTIVE AND THANKS TO DISTRIBUTED REPRESENTATION NO OF PARAMS IS LINEAR IN DIMENSIONS OF h

* LOG-LIKELIHOOD CRITERION INDUCES A LEARNER NOT PERFECTLY GENERALIZING AN ESTIMATOR MUCH SMOOTHER THAN TARGET DISTRIBUTION. BECAUSE ENTROPY *

REGULARIZED AUTOENCODERS

- TRAINING AEs IMPLIES 'BALANCE' BETWEEN TWO FORCES: 1 - LEARNING REPRESENTATION ALLOWING (APPROXIMATE) RECOVERY 2 - CONSTRAINT/REGULARIZATION; BOTTLENECKS, CONTRACTIVENESS, LOG PRIORS...

→ ONLY VARIATIONS NEEDED TO DISCRIMINATE BETWEEN TRAINING EXAMPLES NEED TO BE REPRESENTED **IF DATA AROUND MANIFOLD** → REPRESENTATIONS IMPLICITLY CAPTURE LOCAL COORDINATES FOR MANIFOLD. • **ONLY VARIATIONS TANGENT TO MANIFOLD** AROUND X NEED TO CORRESPOND TO CHANGES IN $h = f(x)$

- ENCODER LEARNS MAPPING ONLY SENSITIVE TO CHANGE ALONG MANIFOLD DIRECTION, NOT TO ORTHOGONAL.

TANGENT DISTANCE

NONPARAMETRIC NN ALGORITHM WHERE METRIC IS DERIVED FROM KNOWLEDGE OF MANIFOLDS → METRIC IS DISTANCE BETWEEN MANIFOLDS / TANGENT ~~POINTS~~ PLANES, BETWEEN TANGENT AND POINT → SOLVABLE WITH A LOW DIMENSIONAL LINEAR SYSTEM

TANGENT-PROP

CLASSIFIER REGULARIZED WITH PENALTY PROMOTING INVARIANCE TO KNOWN LOCAL FACTORS OF VARIATION → SMALL DIRECTIONAL DERIVATIVE $= \lambda \sum_i \left(\frac{\partial f(x)}{\partial x} \cdot v_i \right)^2$
TANGENT VECTORS NEED TO BE KNOWN A PRIORI

MANIFOLD TANGENT CLASSIFIER

LIKE TANGENTPROP, BUT CONTRACTIVE AUTOENCODER TO ESTIMATE THEM BECAUSE THEY LEARN THEM ON THEIR OWN. USE TANGENTS FOR REGULARIZATION