Differential Geometric Regularization for Supervised Learning of Classifiers
Qinxun Bai¹, Steven Rosenberg², Zheng Wu³, Stan Sclaroff¹

¹Computer Science, ²Mathematics & Statistics, Boston University ³The Mathworks Inc.

Motivation
Visual Recognition – Supervised Learning of Classifiers

Four binary and four multiclass datasets in functional space

Counter-intuitive Observations
Dog? Yes + No = =

Rapid Local Oscillation

Smoothness vs. Mean Curvature

Smoothness by functional norms:
- Not specifically tailored to measure local oscillation
- Overskill the hypothesis space
- Sculpting with an axe? Need a sculptor’s knife!

Mean Curvature of the hypersurface:
- In differential geometric sense
- A specific measure of the amount of local oscillation
- Generalizes to high dimensional space

Main Idea
Physical Model

Formal Setup
Learn a function \( f: \mathcal{X} \rightarrow \mathbb{R}^k \) as an estimator of \( P(y|x) \)
The hypersurface associated with \( f \):
\[
\text{graph}(f) = \{(x, f(x), \ldots, f^k(x)) | x \in \mathcal{X} \} \in \mathcal{X} \times \Delta^k
\]

Geometric Foundation on \( \mathcal{H} \)
- Frechet topology on \( \mathcal{H} \), and the induced topology on \( \mathcal{H} \)
- \( \mathcal{H} \) is the volume form of the induced Riemannian metric on graph(f)

RBF Representation

Regularized ERM Formulation
Minimize the regularized loss \( \mathcal{P} \) in functional space \( \mathcal{H} \)

Solve for \( \min_{\mathcal{P}} \mathcal{P}(f) \)

Solve iteratively by gradient flow:
\[
\frac{df_i}{dt} = -\mathcal{P}'(f_i)
\]
- starting from neutral estimator \( f_0 = \left( \frac{1}{n}, \ldots, \frac{1}{n} \right) \)
- evolve \( f_i \) towards \(-\mathcal{P}'(f_i)\)
- \( f_i \) will flow to a local minimum of \( \mathcal{P} \)

Computing \( \mathcal{P}'(f_i) \)
- easy e.g. backpropagation for neural networks

The Gradient \( \mathcal{P}'(f_i) \)

Datasets from UCI Repository
- Four binary and four multiclass datasets
- Following the choice/setup of previous papers

Experiments

RBF Representation

Represent \( f \) as “softmax” output of RBFs
\[
f_i = \exp(k(x, x_i)) / \sum_{j=1}^n \exp(k(x, x_j)), \quad x_i \in \mathcal{X}, \quad i = 1, \ldots, n
\]
where \( k(x) = e^{-|x|^2} \) is the RBF centered at \( x \)

Gradient update for \( A = (a_j) \)
\[
A = A - \varepsilon \mathcal{P}'(f(x_i))\left( f(x_i), \ldots, f(x_n)) \right)^T,
\]
where \( \mathcal{P}'(f(x_i)) = \left( \frac{1}{n} \right)^T \mathcal{P}'(f(x_i)). \quad M_i = \varphi_i(x_i) \)

Real-world datasets – comparing with baseline
- Flicker Material Database (4096 dimensional feature)
- MNIST handwritten digits (60,000 samples)

Mean Accuracy (%)