Manifold-Based Inference for A Supervised Gaussian Process Classifier

Anis Fradi[1,3], Chafik Samir[1,3], and Anne-Françoise Yao[2,3]

[1]CNRS-LIMOS (UMR 6158) & [2]CNRS-LMBP (UMR 6620)

[3]University of Clermont Auvergne, France

Introduction

What?

- A regularization approach, called weighted ridge logistic.
- Two Bayesian approaches, called manifold Laplace approximation (MLA) and manifold expectation propagation (MEP).

Why?

- High dimensionality poses statistical challenges and renders many traditional classification algorithms impractical to use.

How?

- A modified version of ridge logistic, with the advantage of reducing the search space of regularization parameter, which is more efficient for computational time.
- A novel approach having the advantage of learning data into new feature space under some constraints (reduce non-linearity, increase separability) in a Bayesian context.

Discussion and conclusion

- The effectiveness of the proposed approaches has been proved within an application to image classification that contains some defected boxes among correct ones.
- From various conducted tests, we demonstrated that the results are enhanced by both proposed approaches regarding the baseline approach (logistic regression).
- The proposed based on manifold Gaussian process classifier achieve high accuracy.
- The MEP has more predictive power and generalization capability than MLA.
- The difference between MEP and MLA is caused almost exclusively by approximation errors in MLA.

Perspectives

- Estimation of model's hyper-parameters by variational and Bayesian methods.

Regularization approach

- **Summary:** The idea consists of considering the weighted sum between the unstructed log-likelihood of logistic regression and l^2-norm of unknown parameters β.
- **Details:**

Formulation How to find the maximum likelihood estimate (MLE) ?

$$\beta^{\lambda,*} = \arg\max_{\beta} \left\{ l^{\lambda}(\beta) = \frac{(1-\lambda)}{2} l(\beta) - \frac{\lambda}{2}||\beta||_2^2 \right\} \quad ; \quad 0 < \lambda < 1$$

Description of data $\{\mathbb{X}, \mathbb{Y}\} = \{(x_i, y_i)\}_{i=1}^{N}$, where $y_i \in \{0, 1\}$ and $x_i \in \mathbb{R}^p$

Variational methods: Newton/gradient

$$\beta^{k+1} \approx \frac{(1-\lambda)}{2}(-\nabla^2 l^{\lambda}(\beta^k))^{-1}(\nabla^2 l(\beta^k)\beta^k + \nabla l(\beta^k))$$

- **Problems:** $\dim(\beta) = p + 1 >> N \Longrightarrow$ dimensionality reduction.

$\beta^{\lambda,*}$ can be a local maximum $\Longrightarrow$ Bayesian inference gives more candidates through the prior.

Bayesian approaches

- **Summary:** The idea is to learn the Gaussian processes classifier (GPc) in a feature space to get the proposed manifold Gaussian processes classifier (MGPc)
- **Details:**

A MGPc $G : \mathcal{H} \to \mathbb{R}$ is equivalent to a GPc $f = G \circ M : \mathbb{R}^p \to \mathbb{R}$ with a covariance function $\tilde{c}$ such that $\tilde{c}(x, x') = c(M(x), M(x'))$. M is a deterministic and parametrized function obtained by an unsupervised way, that maps the input space $\mathbb{R}^p$ into the feature space $\mathcal{H} \subseteq \mathbb{R}^m$ ($m << p$).

Formulation How to find an explicit form to the posterior proportionality ?

$$\mathbb{P}(\mathbf{G}|\mathbb{Z}, \mathbb{Y}) \propto \mathbb{P}(\mathbf{G}|\mathbb{Z}) \times \prod_{i=1}^{N} \mathbb{P}(y_i|G_i) \quad ; \quad \mathbf{G} = G(\mathbb{Z})$$

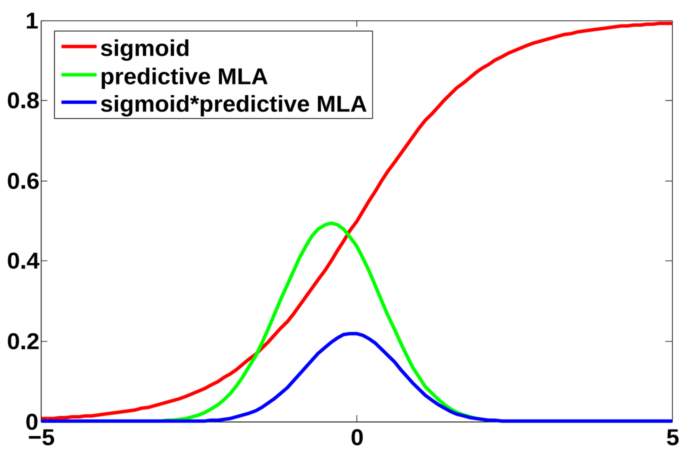
Description of data $\{\mathbb{Z} = M(\mathbb{X}), \mathbb{Y}\} = \{(z_i, y_i)\}_{i=1}^{N}$, where $y_i \in \{-1, 1\}$ and $z_i \in \mathcal{H}$

Approximation methods MLA & MEP

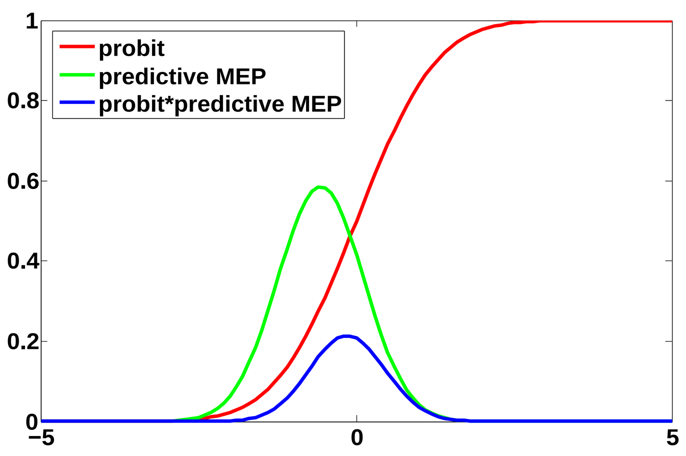
MLA: Employing a Gaussian approximation to the true posterior from the second order Taylor expansion around the MAP estimator.

MEP: Replacing the individual likelihoods by unnormalized Gaussians and minimizing the Kullback-Leiber divergence iteratively between the true posterior and its approximation.

An illustration of key steps to classify a test input ($y^* = -1$)



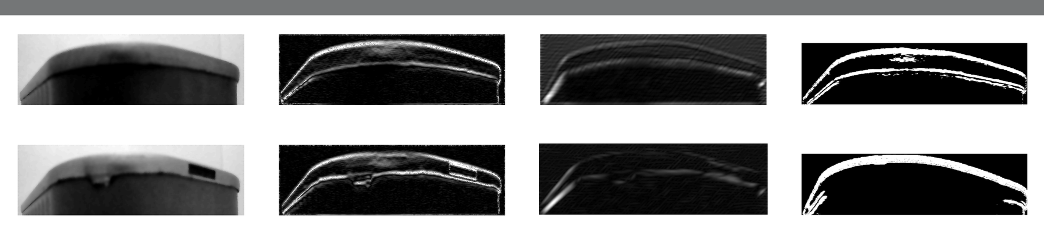
(MLA)



(MEP)

Comment: Approximate predictor of class "1": $\bar{\pi}_* = \mathbf{0.41}$ for MLA & $\bar{\pi}_* = \mathbf{0.33}$ for MEP

Boxes with different representations



Non-defective (top) and defective (bottom) boxes (original, gradient, binarization, Gabor)

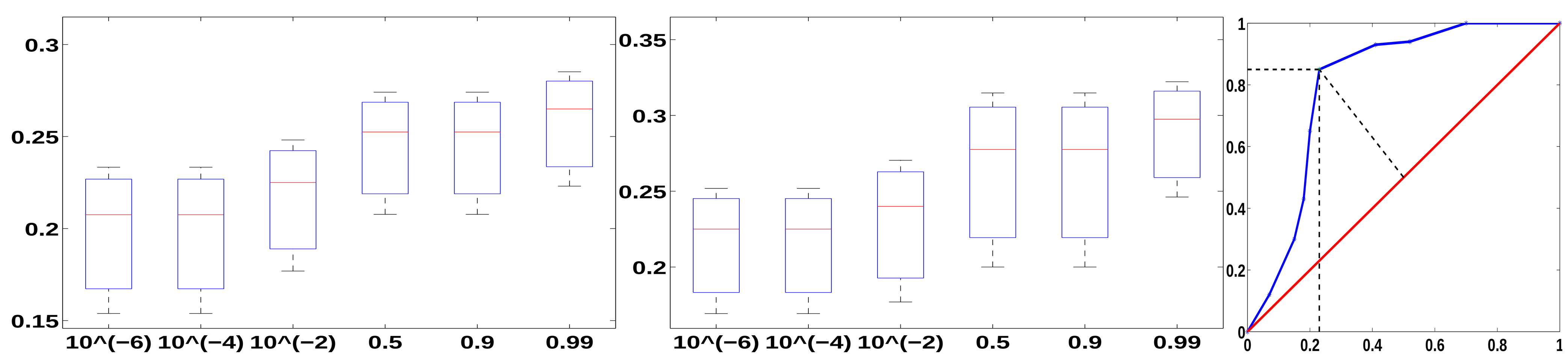
Results: Logistic regression

Error rates \ features	gradient	Gabor	binarization
FP	20%	51%	53%
FN	27%	47%	43%
CE	25%	48%	45%

Results: Bayesian approaches

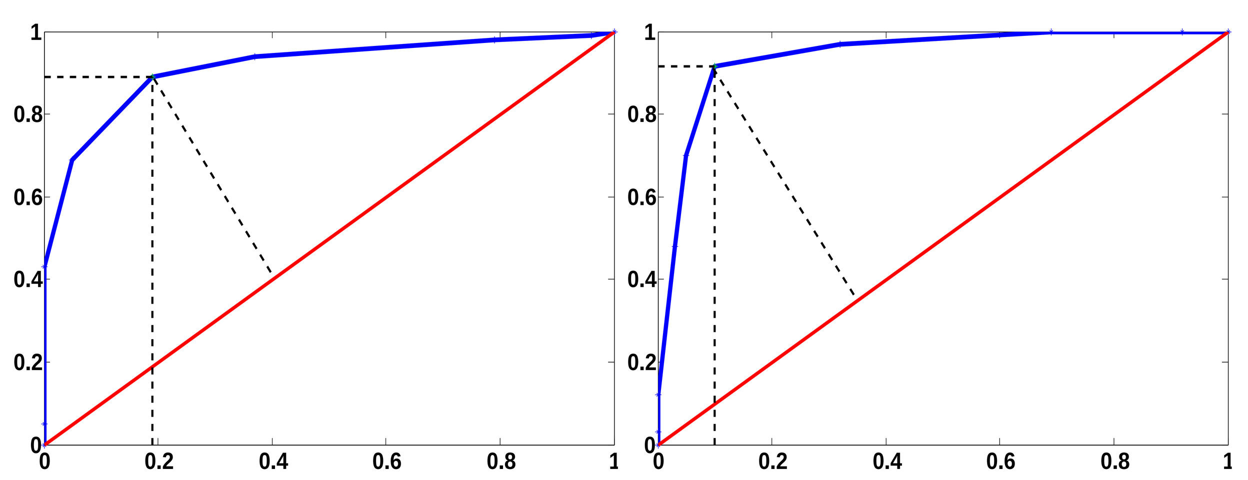
Error rates \ methods	MLA	MEP
FP	11%	8.5%
FN	19%	10%
CE	17%	9.5%

Results: Weighted ridge logistic



Errors as a function of regularization parameters (FN=upper values, FP=lower values and CE=values in the middle) obtained by: Newton (left) and gradient (middle). ROC curve (right)

Results: Figure



ROC curves for Bayesian approaches: MLA (left) and MEP (right)

Anis.FRADI@etu.uca.fr chafik.samir@uca.fr Anne.YAO@uca.fr