Mind the Nuisance: Gaussian Process Classification using Privileged Noise

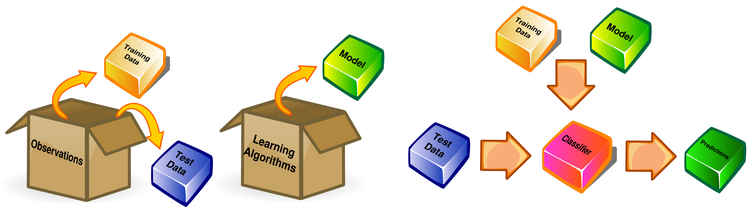
Daniel Hernández-Lobato[1], Viktoriia Sharmanska[2], Kristian Kersting[3],
Christoph H. Lampert[2] and Novi Quadrianto[4]

[1]Universidad Autónoma de Madrid [2]IST Austria [3]TU Dortmund [4]SMiLe CLiNiC, University of Sussex

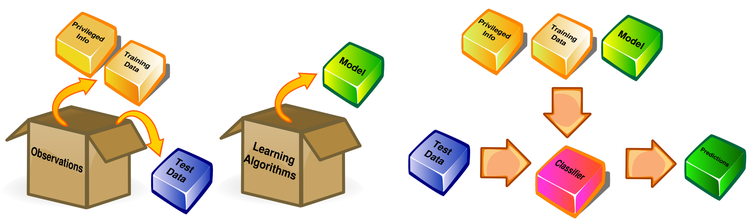
Neural Information Processing Systems Foundation

1. Introduction

Traditional Machine Learning: Given a set of training instances $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ with targets $\mathbf{y} = (y_1, \ldots, y_n)^{\mathrm{T}}$ and a model based on a learning algorithm, we obtain a classifier that can be used for making predictions about new tests instances $\mathbf{x}_{\text{new}}$.



Learning using Privileged Information: Besides $\mathbf{X}$, we have extra information associated to each instance $\mathbf{X}^\star = (\mathbf{x}_1^\star, \ldots, \mathbf{x}_n^\star)$ that is available for learning the classifier. However, this information is only available at training time. This means that it cannot be directly used as an input to the classifier.



2. Examples of Datasets with Privileged Information

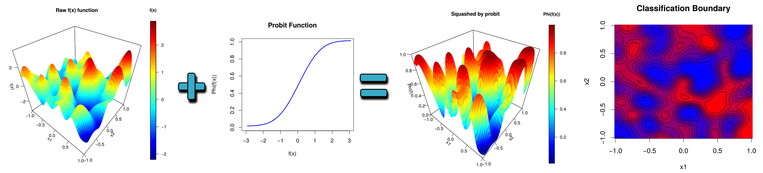
Attribute Discovery Dataset: Contains images of *bags*, *earrings*, *ties* and *shoes*. Each image has associated a textual description.

These classy cultured fresh water pearl earrings are an essential accessory that can be worn everyday as well as on special occasions, with almost anything. Elegantly set in sterling silver, these earrings featuring black pearl will be a favorite in any...

Animals with Attributes Dataset: Besides images of each animal there is extra info such as semantic attributes or DeCAF features obtained from deep networks.

3. Gaussian Process Classification (GPC)

Description: Under this model $p(y_n = 1|\mathbf{x}_n, f) = \Phi_{(0,\sigma^2)}(f(\mathbf{x}_n))$, where σ^2 is the variance of the Gaussian noise around f, and f is assumed to be generated by a Gaussian process, *i.e.*, $f(\mathbf{x}_n) \sim \mathcal{GP}(0, k(\mathbf{x}_n, \cdot))$, for some covariance function $k(\mathbf{x}_n, \cdot)$.



This is equivalent to using $y_n = \text{sign}(\tilde{f}(\mathbf{x}_n)) = \text{sign}(f(\mathbf{x}_n) + \epsilon_n)$, with $\epsilon_n \sim \mathcal{N}(0, \sigma^2)$, for classification. Thus, $\tilde{f}$ is regarded as a nuisance function as we do not observe nor care about its value.

4. Privileged Information inside GPC (GPC+)

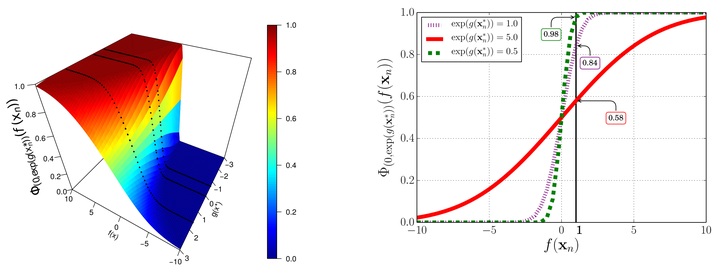
Description: The classification model with privileged noise is:

$$\text{Likelihood model}: p(y_n = 1|\mathbf{x}_n, \tilde{f}) = \mathbf{I}[\,\tilde{f}(\mathbf{x}_n) \geq 0\,], \quad \text{where} \quad \mathbf{x}_n \in \mathbb{R}^d$$

$$\text{Assume}: \tilde{f}(\mathbf{x}_n) = f(\mathbf{x}_n) + \epsilon_n$$

$$\text{Privileged noise}: \epsilon_n \overset{i.i.d.}{\sim} \mathcal{N}(\epsilon_n|0, z(\mathbf{x}_n^*) = \exp(g(\mathbf{x}_n^*))), \quad \text{where} \quad \mathbf{x}_n^* \in \mathbb{R}^{d^*}$$

$$\text{GP prior model}: f(\mathbf{x}_n) \sim \mathcal{GP}(0, k_f(\mathbf{x}_n, \cdot)), \quad g(\mathbf{x}_n^*) \sim \mathcal{GP}(0, k_g(\mathbf{x}_n^*, \cdot)).$$



Privileged information discriminates easy and difficult samples.

5. Expectation Propagation for GPC+

The posterior is approximated by the product of two Gaussians on $\mathbf{f}$ and $\mathbf{g}$:

$$p(\mathbf{f}, \mathbf{g}|\mathbf{y}, \mathbf{X}, \mathbf{X}^\star) = \frac{\prod_{n=1}^N p(\mathbf{y}_n|\mathbf{f}, \mathbf{g}, \mathbf{x}_n, \mathbf{x}_n^\star) p(\mathbf{f}) p(\mathbf{g})}{p(\mathbf{y}|\mathbf{X}, \mathbf{X}^\star)} \approx \mathcal{N}(\mathbf{f}|\mathbf{m}_f, \Sigma_f)\mathcal{N}(\mathbf{g}|\mathbf{m}_g, \Sigma_g)\,.$$

Each factor $p(y_n|\mathbf{x}_n, \mathbf{x}_n^\star, f, g) = \Phi_{(0,\exp(g(\mathbf{x}_n^\star)))}(y_n f(\mathbf{x}_n))$ is approximated as:

$$p(y_n|\mathbf{x}_n, \mathbf{x}_n^\star, f, g) \approx \overline{\gamma}_n(f, g) = \overline{z}_n \mathcal{N}(f(\mathbf{x}_n)|\overline{m}_f, \overline{v}_f)\mathcal{N}(g(\mathbf{x}_n^\star)|\overline{m}_g, \overline{v}_g)\,.$$

The parameters $\overline{z}_n$, $\overline{m}_f$, $\overline{m}_g$, $\overline{v}_f$ and $\overline{v}_g$ can be obtained from the log of:

$$Z_n = \int \Phi_{(0,1)}\left(y_n m_f / \sqrt{v_f + \exp(g(\mathbf{x}_n^\star))}\right) \mathcal{N}(g(\mathbf{x}_n^\star)|m_g, v_g)\, dg(\mathbf{x}_n^\star)\,,$$

and its derivatives which can be approximated using one dimensional quadrature.

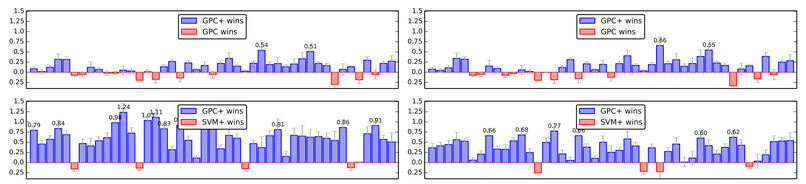
6. Experiments

We compare results with SVM and SVM+. SVM+ is a classifier that uses privileged data to predict the slack variables of the SVM.

Attribute Discovery Dataset:

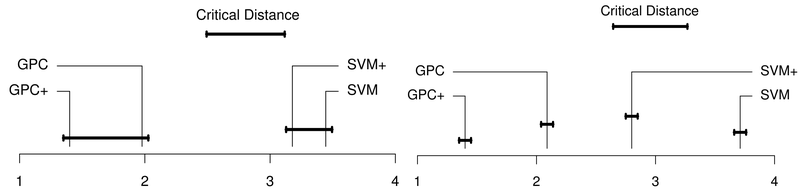
	GPC	GPC+ (Ours)	SVM	SVM+
bags v. earrings	9.79±0.12	**9.50±0.11**	9.89±0.14	9.89±0.13
bags v. ties	10.36±0.16	10.03±0.15	**9.44±0.16**	9.47±0.13
bags v. shoes	9.66±0.13	**9.22±0.11**	9.31±0.12	9.29±0.14
earrings v. ties	10.84±0.14	**10.56±0.13**	11.15±0.16	11.11±0.16
earrings v. shoes	7.74±0.11	**7.33±0.10**	7.75±0.13	7.63±0.13
ties v. shoes	15.51±0.16	15.54±0.16	**14.90±0.21**	15.10±0.18
average error	10.65±0.11	**10.36±0.12**	10.41±0.11	10.42±0.11
average ranking	3.0	**1.8**	2.7	2.5

Animals with Attributes:



(Attributes as privileged) (DeCAF as privileged)

Comparison via the relative difference (length of each bar) between error rates for 45 different cases (top: GPC+ versus GPC, bottom: GPC+ versus SVM+).



(Attributes as privileged) (DeCAF as privileged)

Average rank (the lower the better) of the four methods and critical distance for statistically significant differences (p-value $< 10\%$).

6. Conclusions

- We presented the first treatment of the learning with privileged information paradigm under the GPC framework and called it GPC+.
- In GPC+ privileged information is used in the latent noise layer, resulting in a data-dependent modulation of the slope of the probit likelihood.
- GPC+ is an effective way to use privileged information, which manifest itself in better prediction accuracy at the cost of training two Gaussian processes.
- Recent advances in word-vector neural network representations and deep convolutional networks for image representation can be used as privileged information.

http://dhnzl.org/ http://smileclinic.alwaysdata.net/ daniel.hernandez@uam.es vsharman@ist.ac.at n.quadrianto@sussex.ac.uk