Highlights of Spanish Astrophysics VII, Proceedings of the X Scientific Meeting of the Spanish Astronomical Society held on July 9 - 13, 2012, in Valencia, Spain. J. C. Guirado, L.M. Lara, V. Quilis, and J. Gorgas (eds.)

Signal detection for spectroscopy and polarimetry

A. Asensio Ramos[1,2] **and R. Manso Sainz**[1,2]

[1] Instituto de Astrofísica de Canarias, La Laguna, Tenerife, E-38205 Spain
[2] Departamento de Astrofísica, Universidad de La Laguna, La Laguna, Tenerife, E-38205 Spain

Abstract

Spectroscopic and spectropolarimetric observations with high spectral resolution provide extremely rich information on the physical conditions of distant celestial objects; sometimes, even the mere presence of a spectroscopic or polarimetric pattern may offer fundamental insights. But these are photon-starving techniques. Signals are often at the noise level or buried in it and, many times, just detection proves difficult. Here we present a Bayesian technique for the detection of spectropolarimetric signals based on the application of the non-parametric relevance vector machine to the observations, which allows computing the evidence for the presence of a signal and its most probable value.

1 Introduction

Spectroscopy and spectropolarimetry are two of the most important techniques in the observational astrophysics toolbox. By recording the intensity and polarization state of light at each wavelength we get a quite complete characterization of the state of the light from the observed object, and from its analysis we may infer all the available information on the chemical, thermodynamical, and magnetic properties of the plasma that emitted that light. The main drawback of spectroscopy and spectropolarimetry is that they are often photon-starving techniques. Spectroscopic observations are characterized by the spectral resolution of the spectrograph $R = \lambda/\Delta\lambda$ ($\Delta\lambda$ is the wavelength interval within a resolution element observed at the wavelength λ) which, in the optical and infrared, may typically range $R \sim 1000 - 1000000$ (for low-resolution night-time spectrographs or solar spectrographs, respectively). On the other hand, the fraction of polarized photons P in a light beam is $P \sim$ 1-10% for strongly polarized sources and, typically, $P \lesssim 10^{-3}$. Even worse, polarization is subject to cancellations and P decreases rapidly for low resolution observations. As a consequence, even with the largest telescopes and the most efficient instrumentation the number of (polarized) photons finally reaching a resolution element of the detector may be

very low and close to the noise levels (either the photon noise or the noise of the detection devices), rendering the detection of the signal difficult.

In this contribution we summarize our recent paper [1], where we show how a Bayesian non-parametric regression method can be used for the extraction of spectroscopic and/or spectropolarimetric signals (or any other one-dimensional signal) from noisy observations. The method is based on relevance vector machines [3], a Bayesian version of the support vector machine machine learning technique. Several fundamental advantages are gained. First, we are able to quantify signal detection by computing the evidence ratio between two models: one that contains the signal of interest plus noise and one in which there is only noise. Second, the complexity of the signal is automatically adapted to the information present in the observations. Observations with low noise will facilitate the inference of minute details in the signal of interest, while very noisy observations will favor simpler (and typically smoother) signals. Finally, we obtain an estimation of the signal, together with error bars. We demonstrate the formalism with its application to synthetic and real data.

2 Bayesian signal detection with non-parametric models

Consider the detection of a spectroscopic signal $I(\lambda)$ (equivalently for spectropolarimetric signals) in an observation perturbed with Gaussian noise with zero mean and variance σ^2. In principle, two possibilities may be contemplated. One, what we term model $\mathcal{M}_1$, that there is indeed a signal on the observations $I(\lambda)$ and that it is corrupted with Gaussian noise; the other, model $\mathcal{M}_0$, that there is not such a signal at all, only Gaussian noise. The two options give the following models for the observed signal:

$$\begin{aligned} d(\lambda_i) &= I(\lambda_i) + \epsilon_i, \\ d(\lambda_i) &= \epsilon_i, \end{aligned} \tag{1}$$

where we make explicit that the observed signal is sampled at a set of wavelength points $\{\lambda_i\}_{i=1}^N$. The correct Bayesian way to proceed in order to test for the presence of the signal on given observation (that we represent by the vector $\mathbf{d}$, built by stacking the observed fluxes at all observed wavelength points) is to compute evidence ratio:

$$R = \frac{p(\mathbf{d}|\mathcal{M}_1)}{p(\mathbf{d}|\mathcal{M}_0)}, \tag{2}$$

where the evidence $p(\mathbf{d}|\mathcal{M}_1)$ is the area below the posterior distribution.

Non-parametric regression relies on the application of a sufficiently general function that depends only on observed quantities and that is used to approximate the observations. In our case, the general function is just a linear combination of kernels:

$$I(\lambda; \mathbf{w}) = \sum_{j=1}^{M} w_j K_j(\lambda), \tag{3}$$

where the $K_j(\lambda)$ functions are arbitrary and defined in advance and w_j is the weight associated to the j-th kernel function. In principle, the number of basis functions that one can include into the linear regression can be arbitrarily large (even potentially infinite, in some cases).

The linear regression problem is usually solved by computing the least-squares value of the weights w_j. However, it is known that the least-squares solution leads to severe overfitting and renders the method useless. We considered instead a hierarchical Bayesian solution to the linear regression problem in which the prior for $\mathbf{w}$ is made to depend on a set of hyperparameters $\boldsymbol{\alpha}$, which are learnt from the data during the inference process. The final posterior distribution is then, after following the standard procedure in Bayesian statistics of including a prior for the newly defined random variables, given by:

$$p(\mathbf{w}, \boldsymbol{\alpha}, \sigma^2|\mathbf{d}) = \frac{p(\mathbf{d}|\mathbf{w}, \sigma^2)p(\mathbf{w}, \boldsymbol{\alpha}, \sigma^2)}{p(\mathbf{d})}. \tag{4}$$

Note that the likelihood does depend directly on $\mathbf{w}$ and not on the election of $\boldsymbol{\alpha}$. Assuming that the prior for $\boldsymbol{\alpha}$ and σ^2 are independent and that the prior for $\mathbf{w}$ depend on the hyperparameters $\boldsymbol{\alpha}$, the previous equation can be trivially modified to read:

$$p(\mathbf{w}, \boldsymbol{\alpha}, \sigma^2|\mathbf{d}) = \frac{p(\mathbf{d}|\mathbf{w}, \sigma^2)p(\mathbf{w}|\boldsymbol{\alpha})p(\boldsymbol{\alpha})p(\sigma^2)}{p(\mathbf{d})}. \tag{5}$$

The value of the evidence, or marginal posterior, is computed to ensure that the posterior is normalized to unit hyperarea:

$$p(\mathbf{d}) = \int \mathrm{d}\mathbf{w}\mathrm{d}\boldsymbol{\alpha}\mathrm{d}\sigma^2\, p(\mathbf{d}|\mathbf{w}, \sigma^2)p(\mathbf{w}|\boldsymbol{\alpha})p(\boldsymbol{\alpha})p(\sigma^2), \tag{6}$$

where the priors $p(\mathbf{w}|\boldsymbol{\alpha})$, $p(\boldsymbol{\alpha})$ and $p(\sigma^2)$ are still left undefined.

One of the fundamental ideas of relevance vector machines is to regularize the regression problem by favoring the sparsest solutions, i.e., those that contain the least number of non-zero elements in $\mathbf{w}$. For this reason, and to keep the analytical tractability, it is advantageous to use a product of Gaussian functions for $p(\mathbf{w}|\boldsymbol{\alpha})$:

$$p(\mathbf{w}|\boldsymbol{\alpha}) = \prod_{i=1}^{M} \mathcal{N}(w_i|0, \alpha_i^{-1}), \tag{7}$$

where $\mathcal{N}(w|\mu, \sigma^2)$ is a Gaussian distribution on the variable w with mean μ and variance σ^2. Although not obvious, this prior favors small values of $\mathbf{w}$ when selecting an appropriate prior for $\boldsymbol{\alpha}$. The reason is that, in the hierarchical scheme, the final prior over $\mathbf{w}$ is given by the marginalization:

$$p(\mathbf{w}) = \int \mathrm{d}\boldsymbol{\alpha}\, p(\mathbf{w}|\boldsymbol{\alpha})p(\boldsymbol{\alpha}). \tag{8}$$

If a Jeffreys prior is used for each α_i so that $p(\alpha_i) = \alpha_i^{-1}$, we end up with $p(w_i) \propto |w_i|^{-1}$, which clearly favors small values of w_i. In essence, the form of $p(\mathbf{w}|\boldsymbol{\alpha})$ is such that, in the

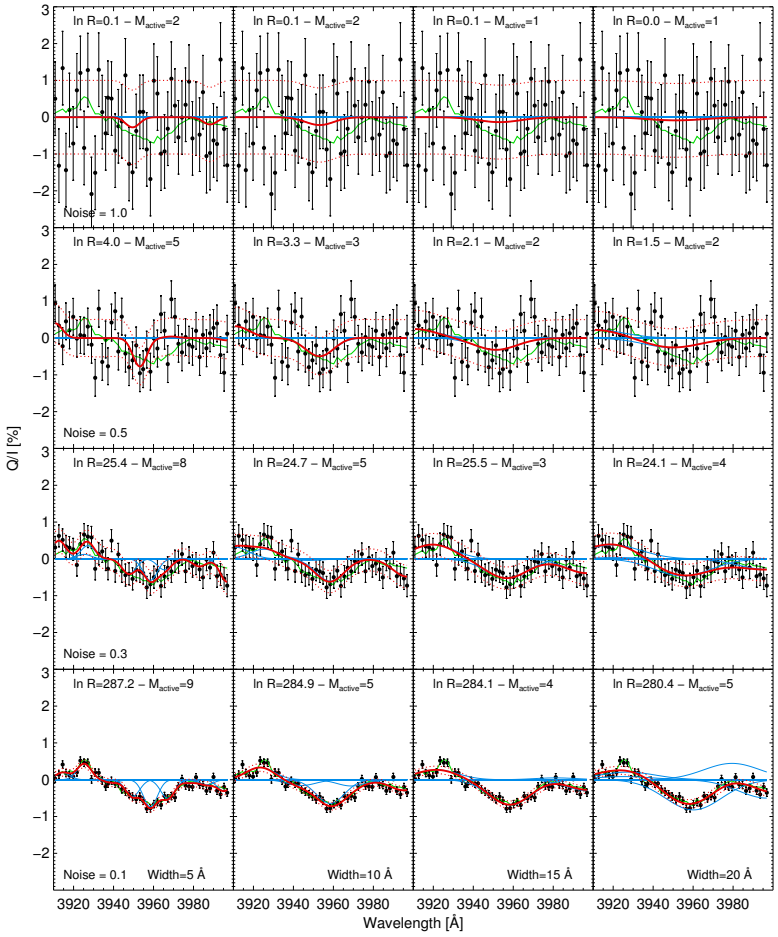

Figure 1: Application to the linear polarization signals in the Ca II H and K lines observed in the atlas of resampled at 50 wavelength points and with different amounts of noise added for each row. The dots display the observations, with their associated Gaussian error bars (with their standard deviation indicated in the panels). Each column shows the results of the line detection using Gaussian functions of different widths as basis functions. The solid red curve is the mean of the predictive distribution, together with the range inside one standard deviation shown in red dotted lines. The blue curves display the contribution of each individual kernel function. Each panel also displays the evidence ratio and the number of active basis functions.

limiting case that α_i tends to infinity, the marginal prior for w_i is so peaked at zero that is compatible with a Dirac delta. This means that this specific w_i does not contribute to the model of Eq. (3) and can be dropped from the model without impact. This regularization proposed by [3] leads to a sparse $\mathbf{w}$ vector, so an automatic relevance determination is implemented in the method.

3 Example

Our example concerns the observation of the linear polarization signals of the H and K lines of Ca II in the UV. These signals have been acquired by [2] at an heliocentric angle of $\mu = \cos\theta = 0.1$ and display an enormous amount of spectral signals that are overlapped with the large-scale structure of the linear polarization of the two Ca II lines produced by quantum superinterferences. We have resampled the profile at a spectral resolution of $\sim$2 Å to mimick a very low spectral resolution spectropolarimeter. The aim is to show that it is possible to detect the linear polarization signal even at such low spectral resolutions under the presence of large noise contaminations.

We have carried out the signal detection procedure for four different levels of Gaussian noise with different standard deviations, as shown in each row of Fig. 1. Given the original (resampled to low resolution) signals (shown in green in the figure), we contaminate them with Gaussian noise so that the S/N in the amplitude peaks of Q/I range from 1 to 10, approximately. The signal detection is done with basis sets composed of Gaussian functions of different widths (each column). The results shown in Fig. 1 look very promising because, even for S/N as low as 1, we can reliably recover the original signal, even though the observed signal is almost unrecognizable. The recovered signal is surprisingly similar to a smoothed version of the green curve, specially when the basis width is large, while many of the minute details of the signal can be estimated correctly if the noise is not too large and the width of the Gaussian basis is small.

Concerning the evidence ratio, we find evidence for signal in all the cases. However, the signal detection algorithm points to a moderate evidence for signal for the case with S/N= 1. The number of active Gaussian functions is usually smaller when the width is larger, with an upper limit of 10 for the smallest considered noise level and width. In any case, we find that the exact green curve is systematically inside one standard deviation of the predictive distribution.

Acknowledgments

Financial support by the Spanish Ministry of Economy and Competitiveness through projects AYA2010-18029 (Solar Magnetism and Astrophysical Spectropolarimetry) and Consolider-Ingenio 2010 CSD2009-00038 is gratefully acknowledged. AAR also acknowledges financial support through the Ramón y Cajal fellowship. This research has benefited from discussions that were held at the International Space Science Institute (ISSI) in Bern (Switzerland) in February 2010 as part of the International Working group *Extracting information from spectropolarimetric observations: comparison of inversion codes*.

References

[1] Asensio Ramos, A. & Manso Sainz, R. 2012, A&A, 547, A113

[2] Gandorfer, A. 2002, *The Second Solar Spectrum*, Vol. II, Zürich: vdf Hochschulverlag

[3] Tipping, M. E. 2000, in *Advances in Neural Information Processing Systems*, 12, 652

[4] Vapnik, V. N. 1995, *The nature of statistical learning theory*, New York: Springer