Pairs Trading Using Gaussian Process Regression: A comprehensive guide

Tanmya Potdar[a,1]

[a]Department of Mathematics

Dr. Puneet Pasricha

Abstract—*This strategy is inspired from the already implemented and tested copula based pairs trading strategy and here we are restricting ourselves to Gaussian distribution although in the future we would with different distributions as well. Pairs Trading in layman's terms is trading on two stocks which show or have shown similar trends in the past. We implement Gaussian Process Regression for probability estimation and then implement a rolling window based trading strategy that although is a very basic strategy but gives positive results indicating that testing this idea over robust strategies might prove beneficial.*

keywords—*Gaussian Process, Bayesian Learning, Pairs Trading*

1. Introduction

One can think of a Gaussian process as defining a distribution over functions, and inference taking place directly in the space of functions, this is also known as the function-space view. But before diving into the function space view or how the process is working we will see a Bayesian Inference based understanding of the same also referred to as the weight space view in existing literature.

1.1. Weight - Space View

We will first understand Bayesian analysis of the Linear Regression Model with **Gaussian Noise**. What we have is

$$f(x) = x^T w \text{ where } \epsilon \sim N(0, \sigma_n^2)$$

This noise assumption gives rise to the *likelihood*, the probability density of the observations given the parameters:

$$p(y|\mathbf{X},\mathbf{w}) = \prod_{i=1}^{n} \mathrm{p}(\mathrm{y}_i|x_i, w) = \mathrm{N}(\mathbf{X}^T w, \sigma_n^2 I)$$

In the Bayesian formalism we need to specify a prior over the parameters, expressing our beliefs about the parameters before we look at the observations. We put a zero mean Gaussian prior with covariance matrix $\sum_p$ on the weights. We then compute the Posterior Distribution i.e.

$$p(\mathbf{w}|\mathbf{X}, y) = \frac{p(y|\mathbf{X},\mathbf{w})p(\mathbf{w})}{p(y|\mathbf{X})}$$

which comes to be via some basic but lengthy calculations as:

$$p(\mathbf{w}|\mathbf{X}, y) = N(\tfrac{1}{\sigma_n^2} A^{-1} X y, A^{-1})$$

$$\text{where } A^{-1} = \tfrac{1}{\sigma_n^2} X X^T + \textstyle\sum_p^{-1}$$

2. Predictive Distribution:

2.1. Test Predictions

To make predictions for a test case we average over all possible parameter values, weighted by their posterior probability. This is in contrast to non Bayesian schemes, where a single parameter is typically chosen by some criterion.

We get the following formula which not so surprisingly is Gaussian again

$$p(f_*|x_*, X, y) = \int p(f_*|x_*, w) p(w|X, y) dw$$

2.2. Projections in Feature Space

The formula we saw above still suffers from limited expressiveness as we are still in the linear domain. A very simple idea to overcome this problem is to first project the inputs into some higher dimension and then applying the linear model there. This projection is done with the help of **kernels** interpreting the same formula but instead of X we will now have the formula corresponding to the kernel trick i.e

$$f_*|X_*, X, f \sim \mathcal{N}(\mu, \sigma^2)$$
$$\text{where}$$
$$\mu = K(X_*, X)K(X, X)^{-1} f$$
$$\textit{and}$$
$$\sigma^2 = K(X_*, X_*) - K(X_*, X)K(X, X)^{-1}K(X, X_*))$$

2.3. Function Space View

Now that we have the predictive distribution we can estimate any future event probabilities given the fact that we know the Kernel and its parameters involved. The first step here is we assume that the mean for the Gaussian Process; but it has also been proved in literature that without considering zero mean we may still get the same results.

2.4. In Context of Trading

Suppose I have the prices of a particular stock for 10 days (training period) then if I treat the price at day t - 1 as my feature and the price at day t as my label, then using the above formula if we are asked to predict the price tomorrow given the price today we can easily do so as we know it is normally distributed ; so now are main goal has reduced to estimating the covariance matrix for the normal distribution which in turn is equivalent to finding the Kernel parameters and this is achieved using **Maximum Likelihood Distribution**. Once have the optimal parameter values are done.

2.5. Trading Strategy

In context of trading on pairs what we do is that once we have the distribution ready with us we compute the conditional probabilities of $U|V$ and $V|U$ where U and V represent the Random Variables associated with price for stock 1 and stock 2 respectively and based on these conditionals we apply the same strategy as we did earlier for the copula based implementation The conditionals are given by the following formula:

$$f_{U|V} \sim \mathcal{N}(\textstyle\sum_{12}^T \sum_{22}^{-1} v,\ \sum_{11} - \sum_{12}^T \sum_{22}^{-1} \sum_{21})$$

2.6. A note on Kernels:

It is important to note that we may use various different kind of kernels and also their combinations to perform the process as demonstrated. The kernels we have used are namely the Radial Basis Kernel, The Rational Quadratic Kernel and the White Noise Kernel.

$$k(x_a, x_b) = \sigma^2 \exp\left(-\frac{||x_a - x_b||^2}{2\ell^2}\right)$$

Figure 1. *Radial Basis Kernel*

$$k(x_a, x_b) = \sigma^2 \left(1 + \frac{||x_a - x_b||^2}{2\alpha\ell^2}\right)^{-\alpha}$$

Figure 2. *Rational Quadratic Kernel*

$$k(x, x) = \sigma^2 I_n$$

Figure 3. *White Noise Kernel*