Proposal

Predicting UK Inflation using a Gaussian Process

Introduction

1. Why I have chosen to model inflation
2. Why I have chosen to use a Gaussian Process Model
    - Compare to a simplistic non-stochastic model

Literature Review

Use primary, *high-quality* sources 1. How and why does GP work * How GP produces a distribution over all the functions that describe the model * General methodology of how GP works 2. Why is GP useful for this model * GP is useful for describing confidence intervals within the function * GP looks at the optimal function family rather than assuming that a model takes a certain trend (linear, quadratic, etc.) 3. Limitations of GP * Cubic run time (may need to do some data cleaning to reduce unnecessary samples and use some efficiency-improving methods) 4. Approximation Methods * Variational Gaussian Process 5. What kernel choices do I have * Kernel choices affect the variance 6. Alteratives to GP (and why're they're not chosen) * Hidden Markov Model for sudden changes in the output (we assume CPI is mostly constant throughout time) * DP for clustering and heirarchical effects in the data (we assume that there's no clustering effect in this dataset) * SVM for non-stochastic regression modelling (we're interested in the stochastic elements of the dataset and consider the CPI to be a random variable) * Bayesian linear regression (linear only form of GP) (we do not assume linearity) 7. Sampling Methods from GP and MVN * Monte Carlo Sampling from a distribution * Variational Gaussian Process (again) 8. Factors of inflation * What are the factors of inflation according to papers * Used to justify variable choice

Methodology

1. How I procured the dataset (using ONS data)
    - Source: ONS
    - Precompilation to CSV file
    - How I read the CSV file to deal with pre-modelling analysis
2. Variable choice in the model
    - Why did I use these variables
3. Statistical Inference methods on variables
    - How did I test for significance
    - How did I test for correlation between variables
4. How I looked at whether the data is clustered
    - Important to see if that we should use something like Gaussian Mixture Models, DP, etc.
5. What methods have I used to improve efficiency and why
    - What are the drawbacks of using these methods
6. Rationalisation on choosing a kernel and its hyper-parameters
    - What kernels did I look at (and why)
    - What methods did I use to decide hyperparameters (Cross Validation, Gradient Descent, etc.)
    - What metrics did I use to decide if the kernel is accurate (MAP on prior)
7. Algorithm for training and testing the GP model
    - How will I build the prior
    - How will I build the posterior
    - How will I test samples from the GP Model
8. Choice on sampling methods (or choices on using variational method)
    - Sampling data to train the model
    - Sampling functions from the GP
9. Statistical inference methods on the prior
    - Significance tests on the parameters of the prior
    - Likelihood function of prior
    - Confidence interval of the prior
10. Statistical inference methods on the posterior
    - Significance tests on the parameters of the posterior
    - Likelihood function of poseterior
    - Confidence interval of the posterior
11. Mathematical equations used
    - Calculating the posterior
    - Calculating the kernel
    - Lower bounds (if they exist), i.e. for VGP
    - Integrating out latent variables

Analysis

1. Analysis of dataset initially (data cleaning and visualisation and statistic inference)
    - What trend does it look like and why
    - Analysis of inference done on the variables
2. Analysis of kernel's effects
    - Visualising the effects on the prior and posterior
    - Analsysis of the metric used to decide the kernel
    - Analsysis of the kernel's learned hyperparameters
    - Exapnd this to other kernels
3. Analysis of prior
    - Visualisation of the prior's distribution
    - Results from Statistical Inference
4. Analysis of posterior
    - Visualisation of the posterior's distribution
    - Results from Statistical Inference
5. Time series analysis techniques on the year variable
    - Optional, but may be interesting

Conclusion

1. Is GP an effective tool for predicting inflation
    - How did I come to this conclusion and why
2. What further tools can be researched
    - Deep Gaussian Process
    - Using DP and alternative methods
    - Using other sampling methods
    - Using other kernels
3. What other time-indexed growth metrics can we measure with GP
    - GDP
    - Wages