Exercise 1: Bayesian inference

07/03/2023 – 14/03/2023

Group D: Clara Kümpel, Damola Agbelese, Yitong Li

* 1. **Maximum likelihood and overfitting**

Polynomial model of order P:

1. Let

Likelihood:

For a single data point y:

For **y** as a vector of data points:

Log-likelihood:

1. Maximum likelihood (ML)

The ML estimator turns out to be the Least Square Error (LSE) estimator.

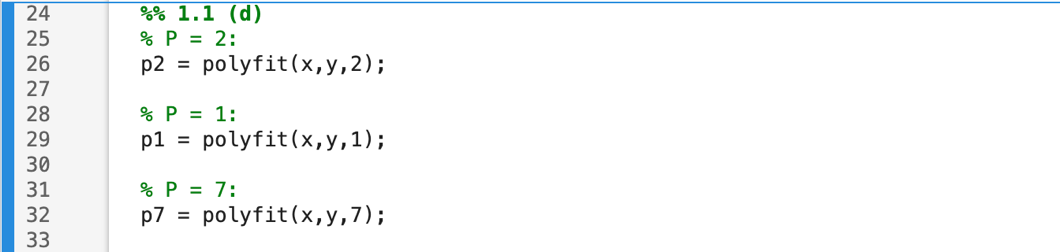
Written with matrices:

Where

Chart

Description automatically generated

1. Considering the analytic solution is the same as the *polyfit* method in MATLAB:



Graphical user interface, text

Description automatically generated

When P = 2, the estimated parameters are much close to the true values.

Text, letter

Description automatically generated

A picture containing application

Description automatically generated

In this case, estimators with higher P result in higher log-likelihood. All estimators perform not much differently.

1. Using new values of x:

Graphical user interface, text, application

Description automatically generated

A picture containing text

Description automatically generated

When fitting a new set of data with parameters generated with the other set of data, performances of estimators with P=1 and P=7 drop drastically, which means they do not generalize well when encountering different data sets.

Performance of estimator with P=2 does not change much.

Chart, histogram

Description automatically generatedChart

Description automatically generatedChart

Description automatically generated with medium confidence

As the order of polynomial term increases, the estimator’s consistency decreases and thus spreads on the histogram across repetitions where different error signals keep being generated.

* 1. **Maximum-A-Posteriori Estimation**

Parameters:

Gaussian prior:

With prior covariance and prior mean

In this exercise,

1. Posterior:

Log-posterior:

1. Maximum-A-Posteriori Estimation:

In our case,

The MAP estimator turns out to be ridge regression (LSE with L2 regularization)

Written with matrices:

Where

Graphical user interface, text

Description automatically generated

Text

Description automatically generated with medium confidenceText

Description automatically generated with medium confidenceText

Description automatically generated with medium confidenceCompared to 1.1 (c), estimators of the first three orders keep relatively invariant, yet deviate from the true value more than 1.1 (c).

The norm of the estimator of P=7 is much smaller than 1.1 (c) where the optimization function does not have the regularization term.

Chart

Description automatically generatedChart

Description automatically generatedChart

Description automatically generated

MAP estimators keep good consistency across repetitions.

* 1. **Bayesian Inference in the Univariate Gaussian Case**

1. Likelihood:

Given 1 fixed data point

1. Posterior of model parameters

Precision:

Define:

We get:

Define:

We get:

Where: