Modern Family Sample Code

Vector Version

Uwe Naumann

Informatik 12:
Software and Tools for Computational Engineering

RWTH Aachen University

Contents

Outline

Objective and Learning Outcomes

Cover Story

Learning from Data

- Model Calibration
- Linear Model
- Nonlinear Model

Summary and Next Steps

Objective

- Introduction to non-scalar Modern Family sample problem and code

Learning Outcomes

- You will understand
  - motivation for the sample problem;
  - mathematical formulation of the sample problem.
- You will be able to
  - build and run the sample code.

Outline

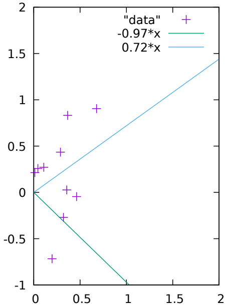
Modern Family Sample Code

Cover Story

© my wife and me

Outline

We aim to understand / reproduce / predict reality (of a given target system, e.g, the Modern Family scenario) based on observed / measured data



$$(X, \mathbf{y}) \in \boldsymbol{R}^{m \times n} \times \boldsymbol{R}^{m}$$

(e.g, positions visited by the guys) through determination of a mathematical model

$$y = f(\mathbf{p}, \mathbf{x}) : \boldsymbol{R}^{n} \times \boldsymbol{R}^{n} \rightarrow \boldsymbol{R}$$

with **x** describing the state of the system (e.g, sets of x-coordinates in between pub and home) and with a vector of (free) model parameters **p** to be determined such that the results obtained from the model match the data in the best possible way.

Our discussion will be restricted to scalar models ($y \in \boldsymbol{R}$). Without loss of generality, we let both **p** and **x** be in $\boldsymbol{R}^n$.

- An exact match with the m data points can be constructed by interpolation, e.g. linear, cubic splines, polynomial of degree m.
- While unknown values in between known data points ("inside" the data) may be predicted adequately this approach typically yields poor extrapolation properties, that is, prediction of values "outside" of the data becomes difficult.
- Moreover, an exact match may not even be desirable due to errors and uncertainty in the given data.
- We chose to select a parameterized mathematical model to reflect the qualitative behavior of the underlying problem. This model is calibrated against the given data to improve the accuracy of its quantitative results.
- Calibration amounts to modification of the values of the (free) model parameters with the objective to minimize some "average" error between the prediction of the model and the given data.

Minima of unconstrained nonlinear convex objectives such as

$$E = E(\mathbf{p}, X, \mathbf{y}) = \sum_{i=0}^{m-1} (f(\mathbf{p}, \mathbf{x}_i) - y_i)^2 \ ,$$

where $X = (\mathbf{x}_i^T)_{i=0}^{m-1}$, $\mathbf{p}, \mathbf{x}_i \in \boldsymbol{R}^n$, are characterized by the following

- necessary (first order) optimality condition:

$$E' \equiv \frac{dE}{d\mathbf{p}}(\mathbf{p}, X, \mathbf{y}) = \mathbf{0} \in \boldsymbol{R}^{1\times n}$$

  (vanishing gradient implies stationary point)

- sufficient (second order) optimality condition:

$$\mathbf{v}^T \cdot \underbrace{\frac{d^2 E}{d\mathbf{p}^2}(\mathbf{p}, X, \mathbf{y})}_{\equiv E'' \in \boldsymbol{R}^{n\times n}} \cdot \mathbf{v} > 0 \quad \forall \mathbf{v} \neq \mathbf{0} \in \boldsymbol{R}^n$$

  (symmetric positive definite (s.p.d.) Hessian implies strict convexity in a neighborhood of the stationary point).

Objective

$$E = \sum_{i=0}^{m-1} (f(\mathbf{p}, \mathbf{x}_i) - y_i)^2$$

Gradient

$$E' = 2 \cdot \sum_{i=0}^{m-1} \left((f(\mathbf{p}, \mathbf{x}_i) - y_i) \cdot \frac{df}{d\mathbf{p}}(\mathbf{p}, \mathbf{x}_i) \right)$$

Hessian

$$E'' = \frac{dE'^T}{d\mathbf{p}} = 2 \cdot \sum_{i=0}^{m-1} \left(\frac{df}{d\mathbf{p}}(\mathbf{p}, \mathbf{x}_i)^T \cdot \frac{df}{d\mathbf{p}}(\mathbf{p}, \mathbf{x}_i) + \frac{d^2 f}{d\mathbf{p}^2}(\mathbf{p}, \mathbf{x}_i) \cdot (f(\mathbf{p}, \mathbf{x}_i) - y_i) \right)$$

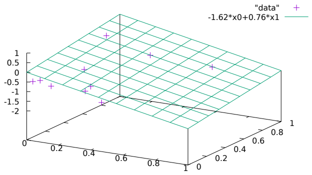
Let the model f be linear in $\mathbf{p}$ satisfying the constraint $f(p,\mathbf{0})=0$, i.e,

$$f=f(\mathbf{p},\mathbf{x})=\mathbf{p}^T\cdot\mathbf{x}=\mathbf{x}^T\cdot\mathbf{p}=\sum_{j=0}^{n-1}x_j\cdot p_j \quad\Rightarrow\quad E=\sum_{i=0}^{m-1}(\mathbf{x}_i^T\cdot\mathbf{p}-y_i)^2$$

First and second derivatives of the objective E are formulated in terms of first and second derivatives of the model f:

$$f'=\frac{df}{d\mathbf{p}}(\mathbf{x})=\mathbf{x}^T \quad\Rightarrow\quad E'=2\cdot\sum_{i=0}^{m-1}(\mathbf{x}_i^T\cdot\mathbf{p}-y_i)\cdot\mathbf{x}_i^T$$

$$f''=\frac{df'^T}{d\mathbf{p}}=\frac{d^2f}{d\mathbf{p}^2}=0 \quad\Rightarrow\quad E''=\frac{dE'^T}{d\mathbf{p}}(X)=2\cdot\sum_{i=0}^{m-1}\mathbf{x}_i\cdot\mathbf{x}_i^T\ .$$



Our sample implementation generates pseudo-random observations plotted as "data" and computes the optimal $\mathbf{p} = (-1.62185, 0.762981)^T$ starting from an initial guess of $\mathbf{p} = (0.0258648, 0.678224)^T$.

$$m = 10, \;\; n = 2$$

At the solution we find

$$E = 3.61678, \quad E' = \begin{pmatrix} 3.30036e-07 \\ -7.62552e-06 \end{pmatrix}, \quad E'' = \begin{pmatrix} 2.34613 & 2 \\ 2 & 2.93244 \end{pmatrix}.$$

The Hessian is s.p.d.

See sample code.

Let the model f be nonlinear in $\mathbf{p}$ satisfying the constraint $f(p, \mathbf{0}) = 0$, e.g,

$$f(\mathbf{p}, \mathbf{x}) = (\mathbf{p}^T \cdot \mathbf{x})^2 = (\mathbf{x}^T \cdot \mathbf{p})^2 = \left(\sum_{j=0}^{n-1} x_j \cdot p_j\right)^2 \quad \Rightarrow \quad E = \sum_{i=0}^{m-1} \left((\mathbf{x}_i^T \cdot \mathbf{p})^2 - y_i\right)^2$$

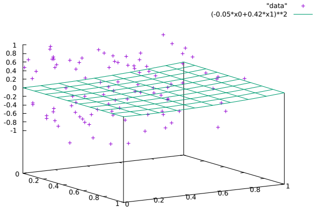
First and second derivatives of the objective E are formulated in terms of first and second derivatives of the model f:

$$f' = \frac{df}{d\mathbf{p}}(\mathbf{p}, \mathbf{x}) = 2 \cdot \mathbf{x}^T \cdot \mathbf{p} \cdot \mathbf{x}^T \quad \Rightarrow \quad E' = 4 \cdot \sum_{i=0}^{m-1} \left((\mathbf{x}_i^T \cdot \mathbf{p})^3 - y_i \cdot \mathbf{x}_i^T \cdot \mathbf{p}\right) \cdot \mathbf{x}_i^T$$

$$f'' = \frac{df'^T}{d\mathbf{p}}(\mathbf{p}, \mathbf{x}) = 2 \cdot \mathbf{x} \cdot \mathbf{x}^T \quad \Rightarrow \quad E'' = \frac{dE'^T}{d\mathbf{p}} = 4 \cdot \sum_{i=0}^{m-1} (3 \cdot (\mathbf{x}_i^T \cdot \mathbf{p})^2 - y_i) \cdot \mathbf{x}_i \cdot \mathbf{x}_i^T \ .$$

Nonlinear Model

Live Demo



Our sample implementation generates pseudo-random observations plotted as “data” and computes the optimal $\mathbf{p} = (-0.0490591, 0.416727)^T$ starting from an initial guess of $\mathbf{p} = (-0.405423, 0.809865)^T$.

$$m = 100, \;\; n = 2$$

At the solution we find

$$E = 30.2784, \quad E' = \begin{pmatrix} -6.01625e-07 \\ 3.45335e-06 \end{pmatrix}, \quad E'' = \begin{pmatrix} 6.45823 & 5.42304 \\ 5.42304 & 16.0209 \end{pmatrix}.$$

The Hessian is s.p.d.

See sample code.

Outline

Summary

- Calibration problem and sample code were introduced.
- Linear and nonlinear models were considered.

Next Steps

- Download the sample code.
- Inspect the sample code
- “Play” with the sample code
- Continue the course to find out more ...