Gaussiset prosessit derivaattahavainnoilla regressio-ongelmassa

(valmiin työn esittely)

Tuomas Nikoskinen

Ohjaaja: TkT Aki Vehtari
Valvoja: Prof. Harri Ehtamo

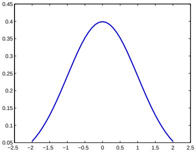
Kandidaattiseminaari 2010
1.11.2010

Esityksen rakenne

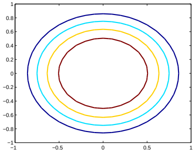
- **Tausta**
- **Derivaattahavaintojen huomiointi**
- **Tavoitteet**
- **Tulokset**
- **Johtopäätökset**

Taustaa: Mitä ovat gaussiset prosessit

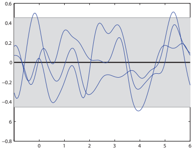
- *Gaussinen prosessi* (GP) on stokastinen prosessi, joka määrää todennäköisyysjakauman yli funktioiden.
- Keskiarvo- ja kovarianssifunktio karakterisoivat gaussisen prosessin täysin.
- Äärellisellä määrällä satunnaismuuttujia kyseessä on moniulotteinen normaalijakauma GP:n keskiarvo- ja kovarianssifunktioilla.



(a) $\mathcal{N}(0,1)$



(b) $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

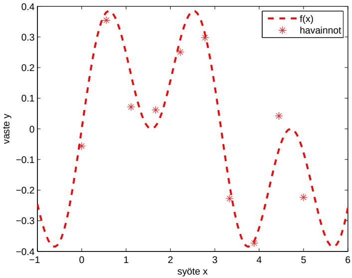


(c) $\mathcal{GP}(m(\mathbf{X}), k(\mathbf{X}))$

Taustaa: Regressio-ongelma & Bayesilainen näkökulma ratkaisuun

- Regressio-ongelma:

$$y(\mathbf{X}) = f(\mathbf{X}) + \epsilon,$$



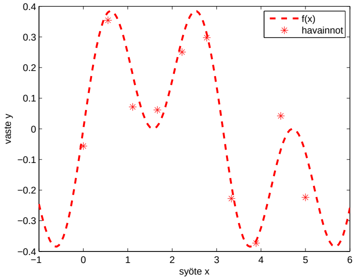
Taustaa: Regressio-ongelma & Bayesilainen näkökulma ratkaisuun

- Regressio-ongelma:

$$y(\mathbf{X}) = f(\mathbf{X}) + \epsilon,$$

- Bayesilaisittain:

$$p(\mathbf{f} \mid \mathbf{y}) = \frac{p(\mathbf{y} \mid \mathbf{f})p(\mathbf{f})}{p(\mathbf{y})}$$

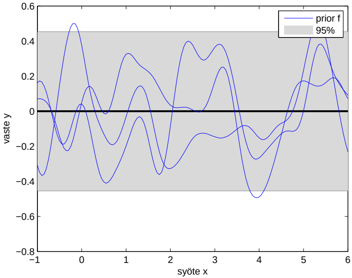


Taustaa: Regressio-ongelma & Bayesilainen näkökulma ratkaisuun

$$p(\mathbf{f} \mid \mathbf{y}) = \frac{p(\mathbf{y} \mid \mathbf{f})p(\mathbf{f})}{p(\mathbf{y})},$$

- Asetetaan prioriksi gaussinen prosessi $f \sim GP$, jolloin äärellisellä määrällä havaintoja

$$p(\mathbf{f}) = \mathcal{N}(\mathbf{0}, \mathbf{K}),$$

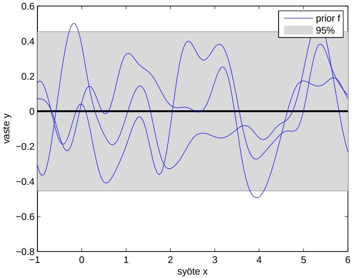


$$p(\mathbf{f}\,|\,\mathbf{y}) = \frac{p(\mathbf{y}\,|\,\mathbf{f})p(\mathbf{f})}{p(\mathbf{y})},$$

- Asetetaan prioriksi gaussinen prosessi $f \sim GP$, jolloin äärellisellä määrällä havaintoja

$$p(\mathbf{f}) = \mathcal{N}(\mathbf{0}, \mathbf{K}),$$

- Priori on tn.jakauma yli funktioiden

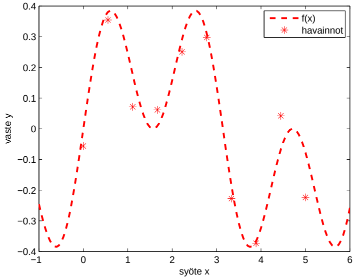


Taustaa: Regressio-ongelma & Bayesilainen näkökulma ratkaisuun

$$p(\mathbf{f} \mid \mathbf{y}) = \frac{p(\mathbf{y} \mid \mathbf{f})p(\mathbf{f})}{p(\mathbf{y})},$$

- Käytetään gaussista uskottavuusfunktiota:

$$p(\mathbf{y} \mid \mathbf{f}) = \mathcal{N}(\mathbf{f}, \epsilon)$$

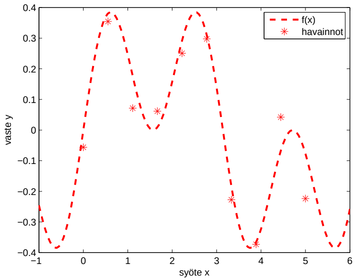


$$p(\mathbf{f} \mid \mathbf{y}) = \frac{p(\mathbf{y} \mid \mathbf{f})p(\mathbf{f})}{p(\mathbf{y})},$$

- Käytetään gaussista uskottavuusfunktiota:

$$p(\mathbf{y} \mid \mathbf{f}) = \mathcal{N}(\mathbf{f}, \epsilon)$$

- $\Rightarrow$ Posteriori laskettavissa analyyttisesti

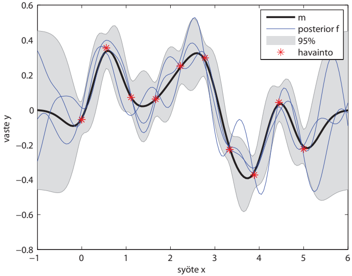


Taustaa: Regressio-ongelma & Bayesilainen näkökulma ratkaisuun

$$p(\mathbf{f} \mid \mathbf{y}) = \frac{p(\mathbf{y} \mid \mathbf{f}) p(\mathbf{f})}{p(\mathbf{y})},$$

- Posteriorista tulee myös gaussinen prosessi

$$p(\mathbf{f} \mid \mathbf{y}) = \mathcal{N}(\mathbf{m}, \mathbf{\Sigma})$$



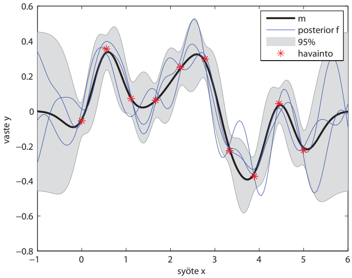
Taustaa: Regressio-ongelma & Bayesilainen näkökulma ratkaisuun

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

- Posteriorista tulee myös gaussinen prosessi

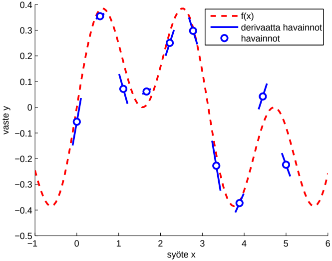
$$p(\mathbf{f} \mid \mathbf{y}) = \mathcal{N}(\mathbf{m}, \boldsymbol{\Sigma})$$

- Jakauman keskiarvo **m** ennusteena mallinnetulle kuvaukselle
- Varianssista $\boldsymbol{\Sigma}_{ii}$ mallin epävarmuus



Derivaattahavaintojen huomiointi

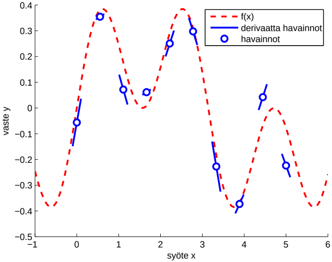
- Kuinkas sitten ne derivaattahavainnot otetaan huomioon?



Derivaattahavaintojen huomiointi

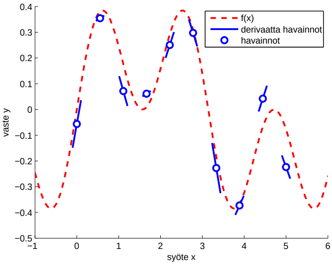
- Kuinkas sitten ne derivaattahavainnot otetaan huomioon?
- Matemaattisesti

$$p(\mathbf{f} \mid \mathbf{y}) \hookrightarrow p(\mathbf{f} \mid \mathbf{y}, \mathbf{y'})$$



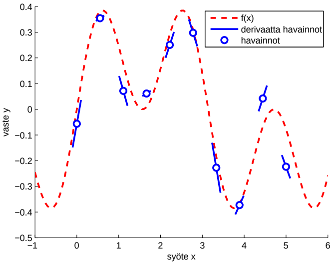
Derivaattahavaintojen huomiointi

- Menettely pääpiirteittäin



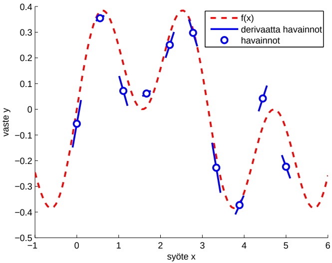
Derivaattahavaintojen huomiointi

- Menettely pääpiirteittäin
  1. Vastehavaintojen jakaumien sijasta tarkastellaan yhteisjakaumia vaste- ja derivaattahavainnoille
     - Derivoitu GP on edelleen GP



Derivaattahavaintojen huomiointi

- Menettely pääpiirteittäin
  1. Vastehavaintojen jakaumien sijasta tarkastellaan yhteisjakaumia vaste- ja derivaattahavainnoille
     - Derivoitu GP on edelleen GP
  2. Ehdollistetaan posteriori sekä vaste- että derivaattahavainnoilla



Tavoitteet

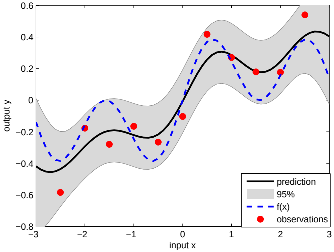
1. Tarkistaa teorian toimivuus
2. Tutkia derivaattahavainnoista saatavaa hyötyä
3. Tarkastella kuinka suuri määrä lisävastehavaintoja kompensoi derivaattahavainnot

Tulokset

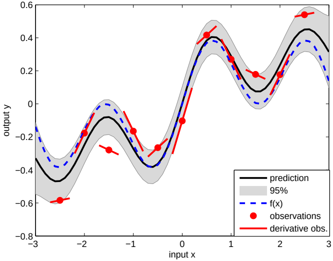
- Derivaattahavaintojen vaikutus epävarmuuteen
  1. Kvalitatiivinen tarkastelu
- Verrataan GP-mallia derivaattahavainnoilla ja ilman regressio-ongelmassa
  2. Yksiulotteinen koetapaus
  3. Kaksiulotteinen koetapaus

Tulokset: Kvalitatiivinen tarkastelu

$$y(x) = \sin(x)\cos(x)^2 + \mathcal{N}(0, 0.15)$$



(d) GP



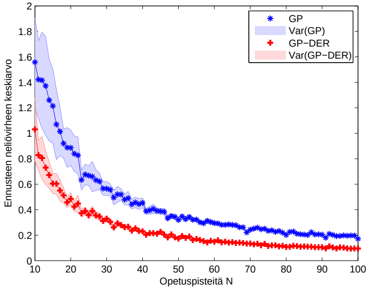
(e) GP-DER

- **Koetapaus 1:** yksiulotteinen regressio-ongelma

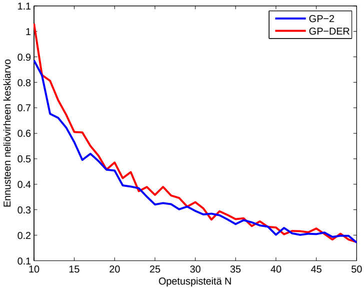
$$y(x) = \sin(x) + \mathcal{N}(0, \sigma^2)$$

$$x \in [-8, 8], \quad \sigma^2 = 0.01$$

Tulokset: Koetapaus 1



(f) GP ja GP-DER mallien neliövirheet



(g) GP-2 mallilla kaksinkertainen määrä havaintoja

- **Koetapaus 2:** kaksiulotteinen regressio-ongelma

$$y(x_1, x_2) = \sin(x_1)\cos(x_2)^2 + \mathcal{N}(0, \sigma^2)$$

$$x_1, x_2 \in [-3, 3], \quad \sigma^2 = 0.01$$

Tulokset: Koetapaus 2

		GP		GP-DER	
Diskr.väli	Havaintoja	tas.	sat.	tas.	sat.
0.60	121	10.83	19.84	3.18	4.20
0.50	**169**	8.08	14.08	**2.21**	2.87
0.40	256	5.63	8.71	1.48	1.72
0.25	625	2.53	3.48	0.62	0.65
0.23	**729**	**2.30**	2.83	0.53	0.59
0.20	961	1.77	2.26	0.38	0.43

Tulokset: Koetapaus 2

		GP		GP-DER	
Diskr.väli	Havaintoja	tas.	sat.	tas.	sat.
0.60	121	10.83	19.84	3.18	4.20
0.50	**169**	8.08	14.08	**2.21**	2.87
0.40	256	5.63	8.71	1.48	1.72
0.25	625	2.53	3.48	0.62	0.65
0.23	**729**	**2.30**	2.83	0.53	0.59
0.20	961	1.77	2.26	0.38	0.43

- GP-DER mallilla $3 \cdot 169 = 507$ havaintoa diskr. välillä 0.5

Tuloksett: Koetapaus 2

		GP		GP-DER	
Diskr.väli	Havaintoja	tas.	sat.	tas.	sat.
0.60	121	10.83	19.84	3.18	4.20
0.50	**169**	8.08	14.08	**2.21**	2.87
0.40	256	5.63	8.71	1.48	1.72
0.25	625	2.53	3.48	0.62	0.65
0.23	**729**	**2.30**	2.83	0.53	0.59
0.20	961	1.77	2.26	0.38	0.43

- GP-DER mallilla $3 \cdot 169 = 507$ havaintoa diskr. välillä 0.5
- $\Rightarrow$ GP-DER malli tarkempi, vaikka havaintoja n. 200 kpl vähemmän!

Johtopäätökset

- Derivaattahavainnot
  - lisäävät merkittävästi GP-mallin tarkkuutta
  - pienentävät GP-mallin epävarmuutta
- Kaksiulotteisessa koetapauksessa derivaattahavainnot arvokkaampia kuin vastaava määrä lisävastehavaintoja
  $\Rightarrow$Derivaattahavaintojen merkitys kasvaa ongelman ulottuvuuksien lisääntyessä?

Kiitos

Tämä työ on tehty Aalto-yliopiston Lääketieteellisen tekniikan ja laskennallisen tieteen laitoksen Laskennallisten kompleksisten systeemien tutkimuksen huippuyksikössä Bayes-tutkimusryhmässä.