S-114.1310 Mallintamisen ja informaatioteorian perusteet

$6 \times 4 = 24$

Tentti
17.12.2008

Ei laskimia!

---

1. Kuvaile Bayesilaisen mallinnuksen kulku mallin määrittelystä päättelyyn ja ennustamiseen kurssin sisällön laajuudessa. Käsittele vastauksessasi myös eroja ja yhtäläisyyksiä perinteiseen frekventistiseen tilastolliseen mallinnukseen. (6p) 5

2. Selitä lyhyesti ja tarkasti seuraavat informaatioteoriaan, tilastolliseen mallintamiseen tai mallinvalintaan liittyvät käsitteet 5

   a) Ennustevirhe (1p)

   b) Epäaito priori (1p)

   c) Estimaattorin harha (1p)

   d) Ristiinvalidointi (1p)

   e) Frekventistinen luottamusväli vs. Bayesilainen luottoväli (1p)

   f) Riippumattomat satunnaismuuttujat (1p)

$18 = 3$

3. Olkoon mitattu näytteet $x_1 = 15$, $x_2 = 21$, $x_3 = 24$, $x_4 = 18$ ja $x_5 = 27$ jostakin prosessista. Tiedetään että mittaukset noudattavat muotoa $x_i = \mu + \epsilon_i$, missä μ on mittauksessa tutkittava suure ja ϵ_i on samalla tavoin riippumattomasti normaalijakautunut mittausvirhe, $\epsilon_i \sim N(0, \sigma^2)$, $\sigma^2 = 25$. Etukäteisoletus tutkittavan suureen käyttäytymisestä on myös normaalijakauma, $\mu \sim N(m_0, s_0^2)$, jossa $m_0 = 25$ ja $s_0^2 = 4$. 4

   a) Muodosta Bayesilainen malli, eli likelihood-funktio ja priorijakauma. Ratkaise tutkittavan suureen μ posteriorijakauma. (3p)

   b) Selitä mitä ovat suurimman uskottavuuden (maximum likelihood, ML) ja maximum a posteriori (MAP) menetelmät parametrien estimoinnissa. Vertaile menetelmien ominaisuuksia ja käyttäytymistä. Ratkaise μ:lle ML- ja MAP-estimaatit. (3p) 4

4. a) Olkoon annettuna yhteisjakauma $p(X, Y)$:

$X \setminus Y$	0	1
0	$\frac{1}{3}$	$\frac{1}{3}$
1	0	$\frac{1}{3}$

   Laske entropiat $H(X)$ ja $H(Y)$, ehdollinen entropia $H(X|Y)$ sekä yhteisinformaatio $I(X; Y)$. (3p)

   b) Mikä on informaatioepäyhtälö (Gibbsin epäyhtälö)? Selitä mitä jakaumien ominaisuutta epäyhtälö kuvaa. (1p)

   c) Mikä on Kraftin epäyhtälö? Selitä miten se liittää prefix-koodit ja todennäköisyysjakaumat toisiinsa. (2p)

$p(y|\varphi, v)p(\varphi)p(v) = p(\varphi|y, v)p(v)$

S-114.1310 Introduction to Modeling and Information Theory

Examination
17.12.2008

No calculators!

---

1. Describe the framework for Bayesian modeling starting from defining the model and ending with inference and prediction in the scope of the course. Also, discuss differences and similarities with traditional frequentist statistical modeling. (6p)

2. Explain in a brief but accurate manner the following terms related to information theory, statistical modeling and model selection

   a) Prediction error (1p)

   b) Improper prior (1p)

   c) Bias of an estimator (1p)

   d) Cross-validation (1p)

   e) Frequentist confidence interval vs. Bayesian credible interval (1p)

   f) Independent random variables (1p)

3. We have measurements $x_1 = 15$, $x_2 = 21$, $x_3 = 24$, $x_4 = 18$ and $x_5 = 27$ from a certain process. It is known that the measurements follow the form $x_i = \mu + \epsilon_i$, where μ is the quantity studied in the measurement and ϵ_i is independently and identically distributed normal noise term, $\epsilon_i \sim N(0, \sigma^2)$, where $\sigma^2 = 25$. Prior assumption about the behavior of the studied quantity is normal distribution, $\mu \sim N(m_0, s_0^2)$, where $m_0 = 25$ and $s_0^2 = 4$.

   a) Formulate the Bayesian model, i.e. likelihood function and prior distribution. Solve the posterior density for the studied quantity μ. (3p)

   b) Explain what is meant by maximum likelihood (ML) and maximum a posteriori (MAP) methods in parameter estimation. Compare the properties and behavior of the methods. Solve the ML and MAP estimates for μ. (3p)

4. a) Let joint probability distribution $p(X, Y)$ be given as

$X \setminus Y$	0	1
0	$\frac{1}{3}$	$\frac{1}{3}$
1	0	$\frac{1}{3}$

   Calculate entropies $H(X)$ ja $H(Y)$, conditional entropy $H(X|Y)$ and mutual information $I(X; Y)$. (3p)

   b) What is information inequality (Gibbs' inequality)? Explain the property of distributions it describes. (1p)

   c) What is Kraft inequality? Explain how it connects prefix codes and probability distributions. (2p)