

UNIVERSITETET I OSLO

Det matematisk-naturvitenskapelige fakultet

Eksamen i: STK4030/STK9030 — Moderne dataanalyse

Eksamensdag: Mandag 13 desember 2010

Tid for eksamen: 14.30 – 18.30

Oppgavesettet er på 2 sider.

Vedlegg: Ingen

Tillatte hjelpemidler: Godkjent kalkulator

Kontroller at oppgavesettet er komplett før du begynner å besvare spørsmålene.

Oppgave 1

- a) La et punkt være uniformt fordelt over enhetskulen i det p -dimensjonale Euklidske rommet \mathbb{R}^p . La Z være avstanden fra origo til dette punktet. Finn fordelingsfunksjonen til Z .

Se på N uavhengige punkter fordelt som ovenfor. La Z_N være punktet nærmest origo. Finn fordelingsfunksjonen til Z_N .

- b) Finn medianen til Z_N .

Hva konkluderer du kvalitativt fra dette når p er stor?

Oppgave 2

- a) Anta sentrerte og skalerte data i design matrisen \mathbf{X} i regresjonsmodellen

$$\mathbf{y} = \beta_0 \mathbf{1} + \mathbf{X}\boldsymbol{\beta} + \mathbf{e},$$

der \mathbf{y} er en p -vektor og \mathbf{X} er $n \times p$.

La $\{\mathbf{v}_m\}$ være egenvektorene til $\mathbf{X}^T \mathbf{X}$, og sett $\mathbf{z}_m = \mathbf{X}\mathbf{v}_m$. Hva kaller du regresjonsmetoden med prediktor

$$\hat{\mathbf{y}}_{(M)} = \bar{y} \mathbf{1} + \sum_{m=1}^M \hat{\theta}_m \mathbf{z}_m \quad \text{med} \quad \hat{\theta}_m = \frac{\mathbf{z}_m^T \mathbf{y}}{\mathbf{z}_m^T \mathbf{z}_m}?$$

Finn en formel for regresjonskoeffisienten for denne metoden.

- b) Vis at metoden gir minste kvadraters regresjon når \mathbf{X} har full rang p og $M = p$.

(Hint: Vis at $\hat{\mathbf{y}}_{(p)} = \bar{y} \mathbf{1} + \mathbf{H}\mathbf{y}$, der \mathbf{H} er hatt-matrisen.)

(Fortsettes på side 2.)

Oppgave 3

- a) Hva er en Bayes klassifikator (Bayes classifier)? Formuler Bayes regel for klassifisering, og fortell hvordan den blir brukt i diskriminantanalyse.
- b) Hva er antakelsene bak lineær diskriminantanalyse? Vis hvordan linearitet blir oppnådd.

Oppgave 4

Hva er en kjerneglatter (kernel smoother) i kurvetilpasning? (Begrens deg selv til det en-dimensjonale tilfellet.) Beskriv to vanlige valg av kjerne.

SLUTT