

UNIVERSITETET I OSLO

Det matematisk-naturvitenskapelige fakultet

Eksamen i: STK4030/9030 — Moderne dataanalyse.

Eksamensdag: Tirsdag 5. desember 2006.

Tid for eksamen: 15.30 – 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.

I en lineær regresjonsmodell har vi

$$E(y|X) = X\beta,$$

der X er en $N \times (p + 1)$ datamatrise.

- Finn minste kvadraters estimator $\hat{\beta}$ for β .
- Vis at $\hat{y} = X\hat{\beta}$ kan skrives som Hy , der H er en projeksjonsmatrise, det vil si en matrise som tilfredsstiller: $H^T = H$ og $H^2 = H$.
- Spesialiser til tilfellet $p = 1$, der første kolonne i X består av 1-tall. Vis at dette gir den vanlige enkle regresjonsmodellen, og vis også at estimatene blir de vanlige for denne situasjonen.

Oppgave 2.

- Hva menes med en kubisk spline?
- I en kubisk spline med K knutepunkter, hvor mange basisfunksjoner trenges? Begrunn svaret.

(Fortsettes side 2.)

Oppgave 3.

Beskriv K -veis (' K -fold') kryssvalidering. Fortell spesielt hvordan den brukes til å estimere prediksjonsfeil.

Oppgave 4.

En trebasert metode tar sikte på å estimere en funksjon

$$f(x) = \sum_{m=1}^M c_M I(x \in R_m).$$

- a) Fortell i grove trekk hvordan regionene R_m bestemmes.
- b) Hvordan estimeres konstantene c_M ? Vis at når to regioner slås sammen, vil det nye estimatet \hat{c}_m ligge mellom de to gamle.

SLUTT