

1. Pohdi Occamin partaveitsen ilmentymiä arkielämässä. Periaatteen mukaan kaikkien muiden seikkojen ollessa yhteneviä, tulee suosia yksinkertaisinta mallia/selitystä. Sama maksimi voitaneen laajentaa myös muille elämäniloille. Tapahuuoku näin? Millaisia (vasta)esimerkkejä löydät?
2. Päätöspuussa ei koskaan testata samaa diskreettiä muuttujaa kahdesti yhdellä juuri-lehti-polulla. Miksi?
3. Oletetaan, että attribuutti jakaa opetusjoukon S osiin S_i , joissa on p_i positiivista ja n_i negatiivista esimerkkiä. Osoita, että attribuutin informaation lisäys on aidosti positiivinen paitsi siinä tapauksessa että suhde $p_i/(p_1 + n_i)$ on sama kaikilla i .
4. Mikä on attribuuttien *Asiakkaita* ja *Tyyppi* informaation lisäys oheisella opetusaineistolla?

Esim.	Attribuutit										Odot
	VE	B	VL	N	Asiakkaita	H	S	V	Tyyppi	Arv.	
x_1	On	Ei	Ei	On	Joitain	\$\$\$	Ei	On	Ranska	0–10	Kyllä
x_2	On	Ei	Ei	On	Täysi	\$	Ei	Ei	Thai	30–60	Ei
x_3	Ei	On	Ei	Ei	Joitain	\$	Ei	Ei	Berger	0–10	Kyllä
x_4	On	Ei	On	On	Täysi	\$	On	Ei	Thai	10–30	Kyllä
x_5	On	Ei	On	Ei	Täysi	\$\$\$	Ei	On	Ranska	>60	Ei
x_6	Ei	On	Ei	On	Joitain	\$\$	On	On	Ital.	0–10	Kyllä
x_7	Ei	On	Ei	Ei	Tyhjä	\$	On	Ei	Burger	0–10	Ei
x_8	Ei	Ei	Ei	On	Joitain	\$\$	On	On	Thai	0–10	Kyllä
x_9	Ei	On	On	Ei	Täysi	\$	On	Ei	Burger	>60	Ei
x_{10}	On	On	On	On	Täysi	\$\$\$	Ei	On	Italia	10–30	Ei
x_{11}	Ei	Ei	Ei	Ei	Tyhjä	\$	Ei	Ei	Thai	0–10	Ei
x_{12}	On	On	On	On	Täysi	\$	Ei	Ei	Burger	30–60	Kyllä

5. Yksinkertainen jälkikarsinta päätöspuille on *reduced error pruning* (Rep), joka opetusjoukon lisäksi vaatii erillisen karsintajoukon esimerkkejä. Annettuna opittu päätöspuu T Rep-algoritmi ensin luokittelee kaikki karsintajoukon tapaukset T :llä pitäen matkalla yllä tarvittavia laskureita ja sen jälkeen alhaalta ylöspäin käy puun läpi karsien siitä alipuut, joiden poistaminen ei kasvata virhettä karsintajoukon luokittelussa. Kirjoita Rep-algoritmi.