1**Tehtävä 1.**
Arvioi kyselyn etäisyyttä tietokannan mittauksista db_1 ja db_2 Euklidisella etäisyydellä, Pearsonin korrelaatiolla ja Spearmanin korrelaatiolla,
kun kysely = (-1, 6, 3) (uusi mittaus) ja
db_1 = (-2, 7, 6)
db_2 = (-1/3, 2, 1).
(Sanallinen tai/ja etäisyyden numeroarvio, myös piirros käy).

Vastaus:
d_eu(q, db_1) = sqrt((-1-(-2))^2+(6-7)^2+(3-6)^2)
d_eu(q, db_1) = 3.316625
d_eu(q, db_2) = sqrt((-1-(-1/3))^2+(6-2)^2+(3-1)^2)
d_eu(q, db_2) = 4.521553
Euklidissa myös koordinaattitarkastelu ja piirros antaa 'selityksen'.

d_pea(q, db_1)=0.94 (alle 1)
> cor((-1, 6, 3),(-2,7,6),method="pearson")
1. ja 3. koordinaatti kaksinkertaistuu (mutta kolmas kuitenkin TÄYSIN eri tavalla kuin db_1:)
d_pea(q, db_2)=1
cor((-1, 6, 3), (-1/3,2,1), method="pearson")
Kun kyselyn koordinaatit kerrotaan 1/3 saadaan db_2 -> täysi korrelaatio

d_spe(q, db_1)=1
> cor((-1, 6, 3),(-2,7,6),method="spearman")
d_spe(q, db_2)=1
> cor((-1, 6, 3),(-1/3, 2, 1),method="spearman")
Arvojen järjestykset kaikissa (1, 3, 2) -> Spearman ei erottele

Tehtävä 2.
Kuvaile, mitä ‘kyselyn etäisyys on nolla’ tarkoittaa SVD-LS-menetelmässä.

Vastaus:
Esimerkkikuvaus, muutkin kelpaa, tämä on vain yksi vaihtoehto, 'piirrokset parhaita' -periaate:
'Kysely on ‘samanlainen’ kuin joku sarakeavaruuden (=mittausten) lineaarikombinaatio'
Ts.: (esimerkiksi)
a) Jäännöksen normi ('pituus') on nolla. -> b)
b) kysely = db_i*x -> c)
c) kysely on db_i:n sarakkeiden lineaarikombinaatio käännöksessä db_i*x, missä x on paras sovitus.
(Käännös merkitsee datamatriisi*rivivektori = sarakevektori)
Jos LS-sovituksessa etsittävä x-vektori sijaitsee
geeniavaruuden kannalla (= lineaarikombinaatio ominaisgeenejä),
se kääntyy riviavaruudesta sarakeavaruuteen (= mittausten lineaarikombinaatio)
yksikäsitteisesti. (Huom. Yleisesti: tietty rivivektori ei käänny tietyksi sarakevektoriksi. Näin käy kuitenkin SVD:n sama-indekseisille kantavektoreille , kun lisäksi skaalataan indeksiltään samalla singulaariarvolla.)

Tehtävä 3.
Lyhyitä ***O*** (oikein) ***V*** (väärin) väittämiä: **merkitse** ***O*** tai ***V***, saa perustella.

1. Jos datamatriisin 1. singulaariarvo on nolla, matriisin joku arvo voi olla muu kuin nolla.

2. Jos 1. ja 2. datamatriisin singulaariarvo ovat kohtalaisia ja muut häviävän pieniä, miittaukset muodostavat tason.

3. Poikkeavat havaintoarvot voi laittaa nollaksi, kun SVD:tä käyttää.

Vastaus: (kaikki vastaukset 'oikeita' ;))
1. ***O*** (laskentatarkkuusraja voi aiheuttaa nollaantumisen, vaikka joku luku olisikin nollasta poikkeava) tai ***V*** (periaatteessa ei)
2. ***O*** (Koska mittaukset sisältävät aina virheitä, näin voidaan olettaa) tai ***V*** (Riippuu mittaustarkkuudesta; häviävän pienikin voi hyvin, hyvin tarkoissa mittauksissa olla merkityksellinen - tämä on kuitenkin erittäin harvinaista)
3. ***O*** (Usen näin on, mutta louhinnassa käytetty 'malli' ja louhintasovelluksen 'vaatimukset' määräävät, onko se järkevää) tai ***V*** (Jos poikkeavia arvoja on paljon, tulee niitä sisältävät rivit tai sarakkeet karsia. Toinen vaihtoehto on 'arvioda (imputation)'/'jopa mitata' data-arvot uusiksi)