

# Sovellusohjelmointi Matlab- ympäristössä

Vertaisverkon koneiden  
klusterointi

# Aiheesta

- Koneiden luokittelu käyttäen kahta Kmeans-toteutusta
- Havaintoavaruuden dimension pienentäminen pääkomponenttianalyysillä
- Kmeans-funktioiden luokittelujen vertailu

# Käytetyt menetelmät: Kmeans

- Jakaa havaintoavaruuden luokkakeskipisteiden eli prototyyppien kesken
- Kuhunkin luokkaan kuuluvat ne pisteet, jotka ovat sitä lähinnä

# Käytetyt menetelmät: PCA

- Tavoitteena vähentää havaintoavaruuden dimensiota niin, että mahdollisimman vähän informaatiota menetetään
- Lasketaan kovarianssimatriisin ominaisarvot ja niitä vastaavat ominaisvektorit
- Valitaan joukko suurimpia ominaisarvoja ja niitä vastaavia ominaisvektoreita → saadaan pääkomponenttien virittämä aliavaruus

# Data-aineisto

- Naapurien lukumäärä,
- koneelta lähtevien kaikkien yhteyksien pienin viive millisekunteina,
- koneen kautta kulkevan resurssikyselyliikenteen määrä kilotavuina,
- lähetettyjen resurssivastausten lukumäärä ja
- tarjolla olevien resurssien lukumäärä.

# Toteutuksesta

- Kmeans:

- toteutus C-kielellä

```
[centres, classes]=kmeans(k, data)
```

Mex-rajapinta:

```
void mexFunction(int nlhs, mxArray  
*plhs[], int nrhs, const mxArray *prhs[])
```

- toisena vertailun kohteena Corneyn Matlab-toteutus

```
[classes, centres, finalDistance]=dcKMeans(da  
ta, k, initCentres, maxIters)
```

# Toteutuksesta

- **PCA**

`[z]=PCA(data)`

- **Distance**

`[dist]=distance(a,b)`

# Toteutuksesta

