



JOHDATUS TEKOÄLYYN

TEEMU ROOS



HELSINGIN YLIOPISTO

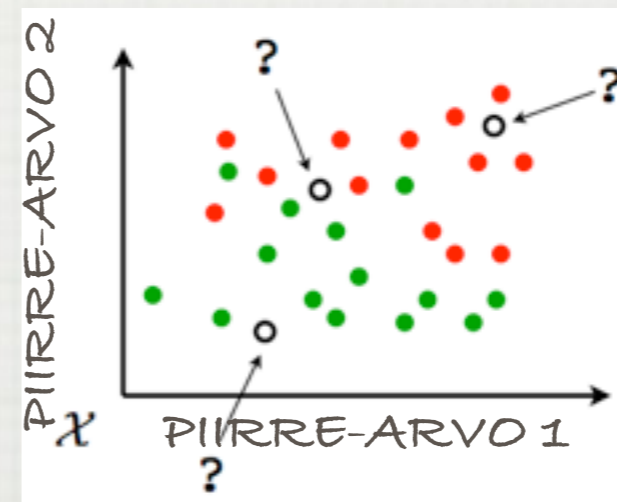
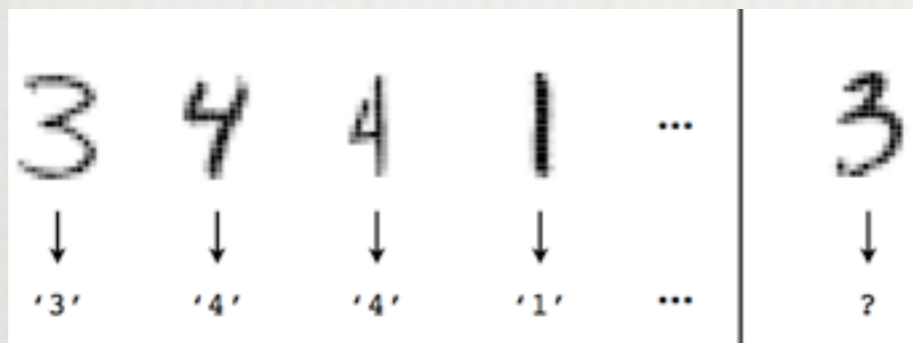
(KALVOT MUOKATTU PATRIK HOYERIN LUENTOMATERIAALISTA)

KONEOPPIMISEN LAJIT

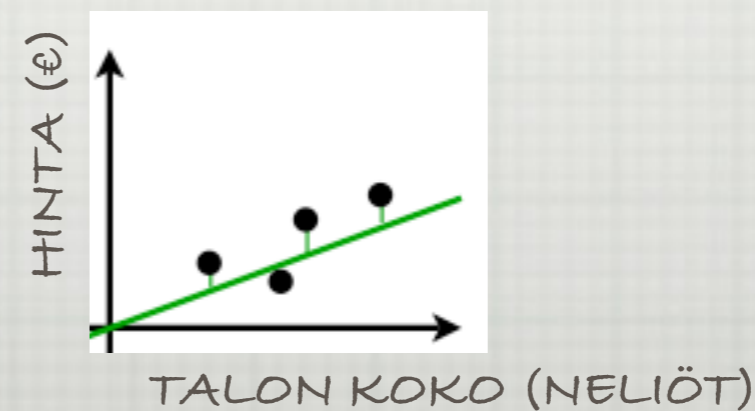
* OHJATTU OPPIMINEN:

- ESIMERKIT OVAT PAREJA (X, Y) , TAVOITTEENA ON OPPIA ENNUSTAMAAN Y ANNETTUNA X .

LUOKITTELU (DISKREETTI 'Y'):



REGRESSIO (JATKUVA 'Y'):

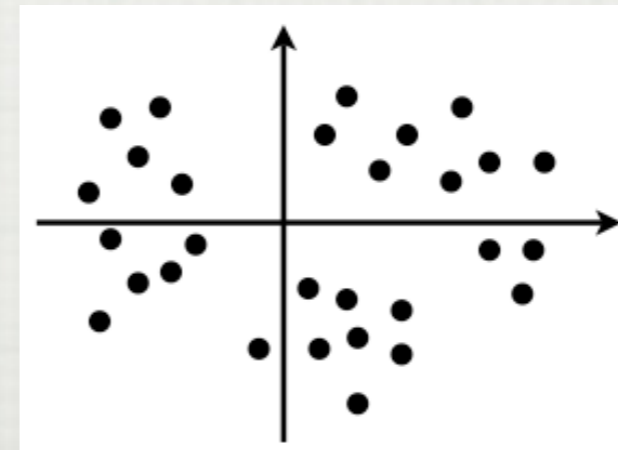
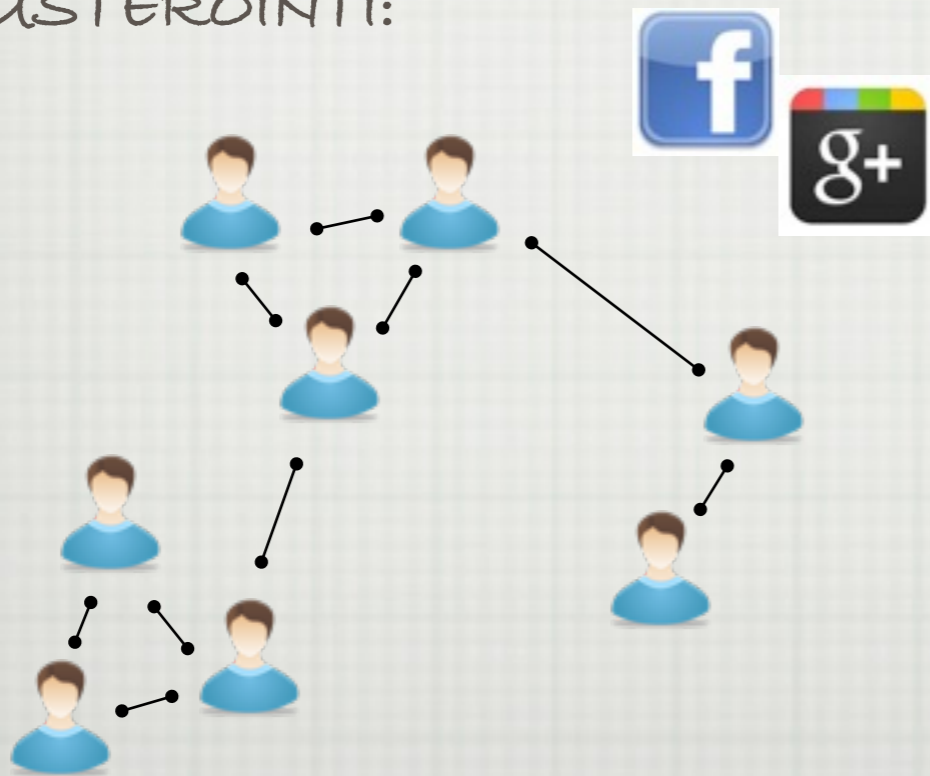


KONEOPPIMISEN LAJIT

* OHJAAMATON OPPIMINEN:

- ESIMERKIT EIVÄT SISÄLLÄ 'OIKEAA VASTAUSTA' Y. SEN SIAAN TAVOITTEENA ON VAIN YMMÄRTÄÄ ANNETTU DATAJOUKKO.

KLUSTEROINTI:

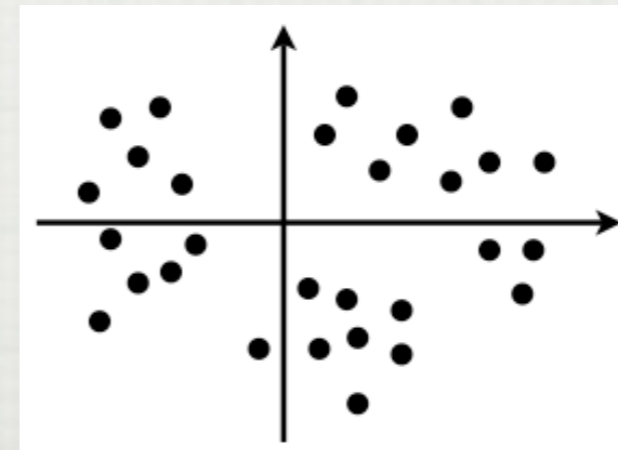
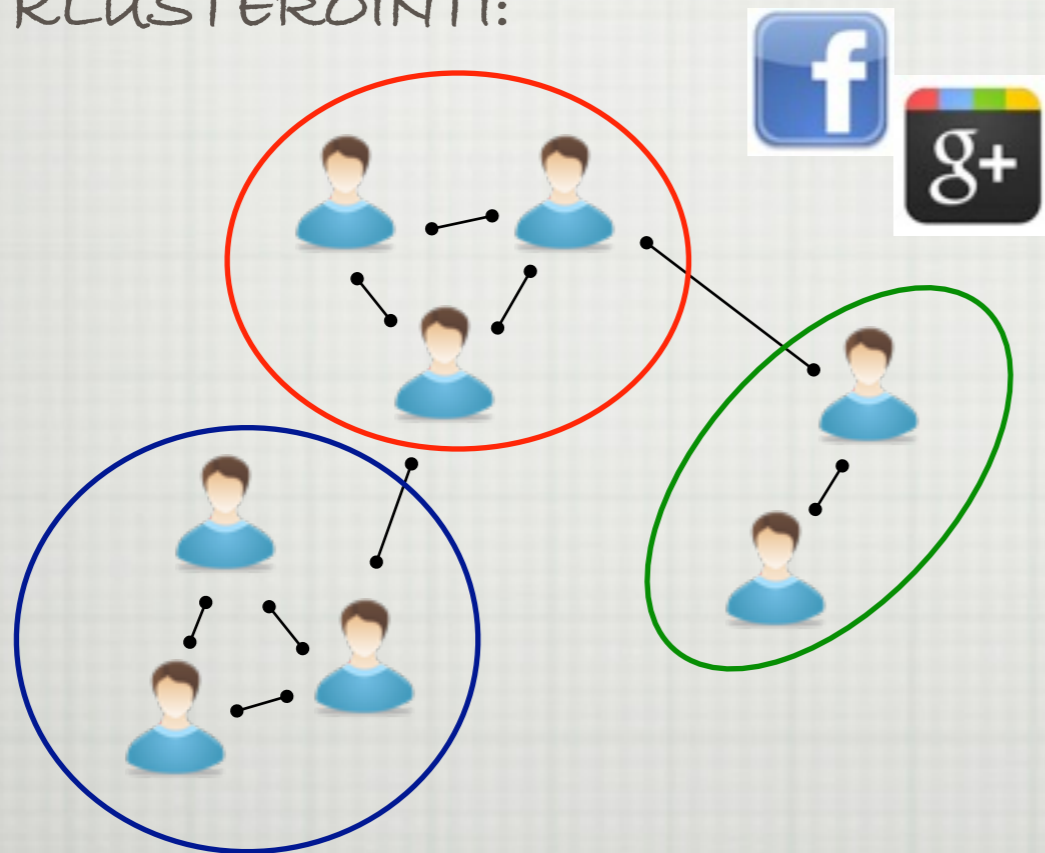


KONEOPPIMISEN LAJIT

* OHJAAMATON OPPIMINEN:

- ESIMERKIT EIVÄT SISÄLLÄ 'OIKEAA VASTAUSTA' Y. SEN SIJAAN TAVOITTEENA ON VAIN YMMÄRTÄÄ ANNETTU DATAJOUKKO.

KLUSTEROINTI:



KONEOPPIMISEN LAJIT



* VAHVISTUSOPPIMINEN:

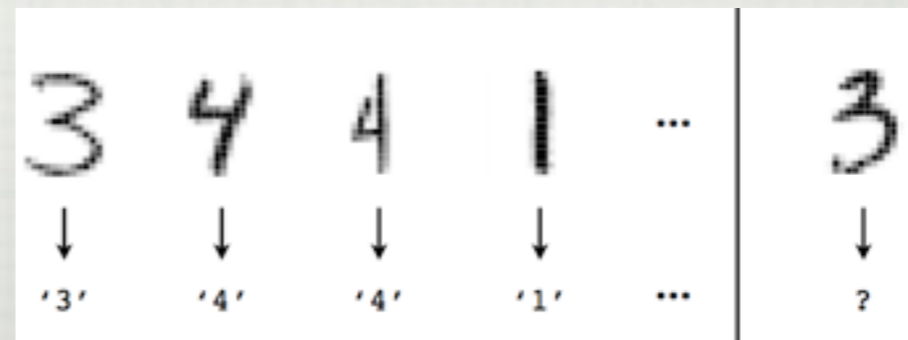
- ANNETTUNA SYÖTE 'X', OHJELMA TULOSTAA 'Y', MUTTA SEN SIAAN ETTÄ MEILLÄ OLISI TARKKA 'OIKEA' Y, OSATAANKIN VAIN ANTAA PALAUTETTA (MAHDOLLISESTI VIIVEELLÄ).
- PALAUTE KERTOO, KUINKA 'HYVÄ' OHJELMA KOKONAISUUDESSAAN OLI
- ELI: KANNUSTETAAN, MUTTEI ANNETA MITÄÄN YHTÄ OIKEAA VASTAUSTA

* KAIKKI KONEOPPIMISTEHTÄVÄT EIVÄT AINAKAAN IHAN SUORAAN MAHDU NÄIDEN YKSINKERTAISTEN OTSIKOIDEN ALLE! (ESIM: 'SEMI-SUPERVISED LEARNING', 'YHTEISÖLLISTÄ SUODATTAMISTA', 'LEARNING TO RANK', ...)

KÄSINKIRJOITETUT MERKIT

KONEOPPIMISONGELMAN MÄÄRITTELY (ESIM.):

- * TEHTÄVÄ: ANNETTUNA KUUSI 28 X 28 PIKSELIKUVA, LUOKITTELE SE JOHONKIN LUOKKAAN 0-9. (FUNKTION SYÖTE ON SIIS $28 \times 28 = 784$ -PITUINEN BINÄÄRIVEKTORI, JA OHJELMAN PITÄÄ PALAUTTAA JOKIN KOKONAISLUKU 0-9.)
- * HYVYYSMITTA: KUINKA MONTA PROSENTTIA KUUSISTA KUVISTA LUOKITELLAAN OIKEIN (ELI SIIHEN LUOKKAAN, JOKA MEIDÄN ESIMERKEISSÄ ON ANNETTUNA)
- * DATA: MNIST-KÄSINKIRJOITETUT NUMEROT, JOSSA JOKAISALLE KUVALLE ON ANNETTU LUOKKA 0-9. ENSIMMÄISET 5000 KPL KÄYTETÄÄN OPETUSDATANA, SEURAAVAT 1000 KPL TESTIDATANA.



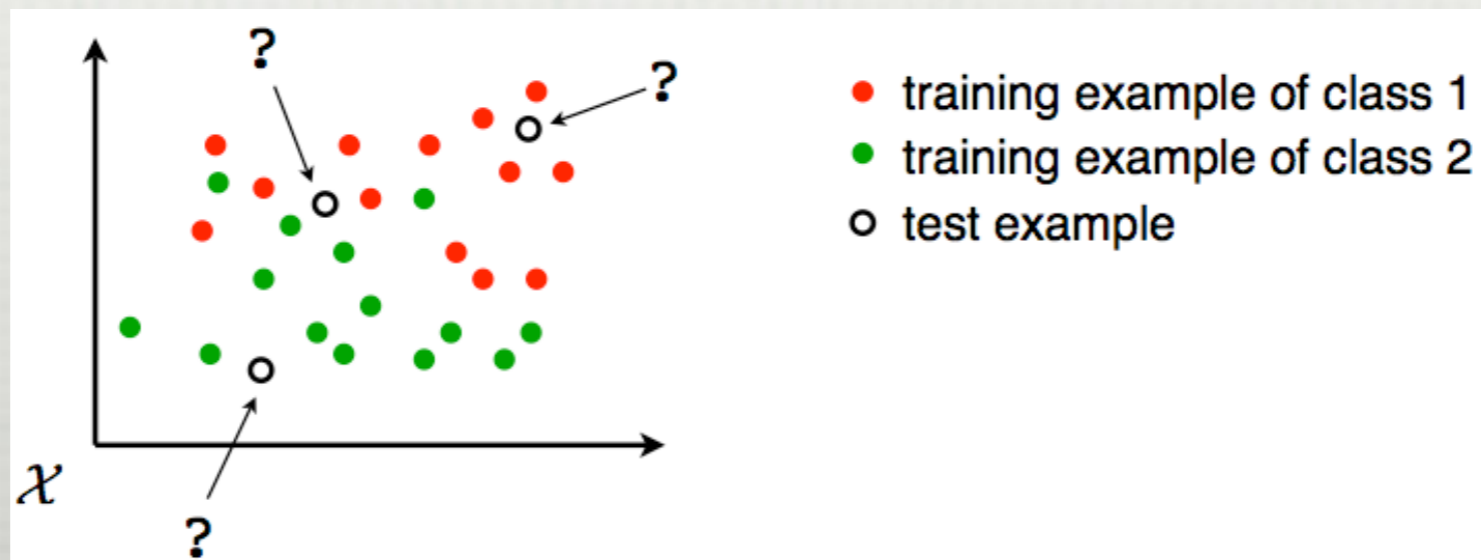
LÄHIMMÄN NAAPURIN LUOKITIN

'YKSINKERTAISIN MAHDOLLINEN' LUOKITTELIJA.

* TOIMINTA:

1. TALLENNA KOKO OPETUSDATA
2. ANNETTUNA UUSI (TESTI-) SYÖTEVEKTORI x , LÖYDÄ SITÄ LÄHIMPÄNÄ OLEVA OPETUSDATAN VEKTORI x^{TRAIN} JA PALAUTA SITÄ VASTAAVA LUOKKA y^{TRAIN} .

* HAVAINNOLLISESTI SIIS NÄIN:



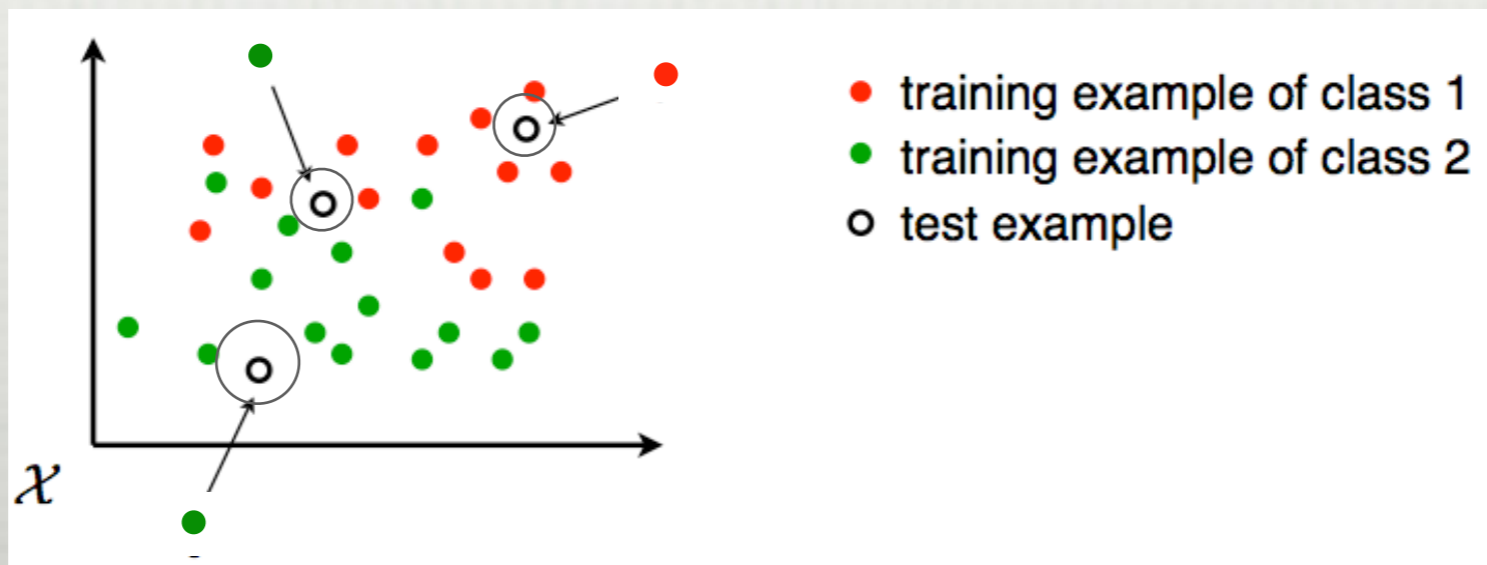
LÄHIMMÄN NAAPURIN LUOKITIN

'YKSINKERTAISIN MAHDOLLINEN' LUOKITTELIJA.

* TOIMINTA:

1. TALLENNA KOKO OPETUSDATA
2. ANNETTUNA UUSI (TESTI-) SYÖTEVEKTORI x , LÖYDÄ SITÄ LÄHIMPÄNÄ OLEVA OPETUSDATAN VEKTORI x^{TRAIN} JA PALAUTA SITÄ VASTAAVA LUOKKA y^{TRAIN} .

* HAVAINNOLLISESTI SIIS NÄIN:



LÄHIMMÄN NAAPURIN LUOKITIN

MERKKIEN VÄLISET ETÄISYYDET?

- * KÄYTÄNNÖSSÄ KAHDEN BINÄÄRIVEKTORIN ETÄISYYTTÄ VOIDAAN MITATA ESIM LASKEMALLA KUINKA MONESSA KOHTAA VEKTORIT OVAT ERISUURET:

1 0 0 0 1 1 0 0 0 1 0 1 0

1 1 0 0 0 1 0 0 1 0 0 1 0

0 1 0 0 1 0 0 0 1 1 0 0 0 ⇒ 4 EROA

...TÄMÄ VASTAA SITÄ ETTÄ LAITETAAN 'MERKIT PÄÄLLEKKÄIN', JA KATSOTAAN KUINKA MONESSA KOHTAA NE EROAVAT.

- * ONKO TÄMÄ HYVÄ ETÄISYYSMITTA? (ESIM KAKSI MELKEIN IDENTTISTÄ MERKKIÄ JOTKA EROAVAT VAIN NIIN ETTÄ TOISESSA KUVASSA MERKKI ON SIIRTYNYT PARIN PIKSELIN VERRAN SAATTAVAT TÄMÄN ETÄISYYSMITAN SUHTEEN OLLA HYVIN KAUKANA TOISISTAAN)

LÄHIMMÄN NAAPURIN LUOKITIN

* TOINEN ESITYS SAMASTA ASIASTA:

OPETUSDATA (LUOKAT ANNETTU):

0 7 1 1 4 9 4 3 4 8 2 2 1 8 7 0 8 1 0 7

0 7 1 1 4 9 4 3 4 8 2 2 1 8 7 0 8 1 0 7

Uudet kuvat:

1
1 7 9 1 1 8 5 7 5 0 6 6 0 4 1 2 3 4 4

LÄHIMMÄN NAAPURIN LUOKITIN

* TOINEN ESITYS SAMASTA ASIASTA:

OPETUSDATA (LUOKAT ANNETTU):

0 7 1 1 4 9 4 3 4 8 2 2 1 8 7 0 8 1 0 7

0 7 1 1 4 9 4 3 4 8 2 2 1 8 7 0 8 1 0 7

Uudet kuvat:

1
1 7 9 1 1 8 5 7 5 0 6 6 0 4 1 2 3 4 4

LÄHIMMÄN NAAPURIN LUOKITIN

* TOINEN ESITYS SAMASTA ASIASTA:

OPETUSDATA (LUOKAT ANNETTU):

0 7 1 1 4 9 4 3 4 8 2 2 1 8 7 0 8 1 0 7

0 7 1 1 4 9 4 3 4 8 2 2 1 8 7 0 8 1 0 7

Uudet kuvat:

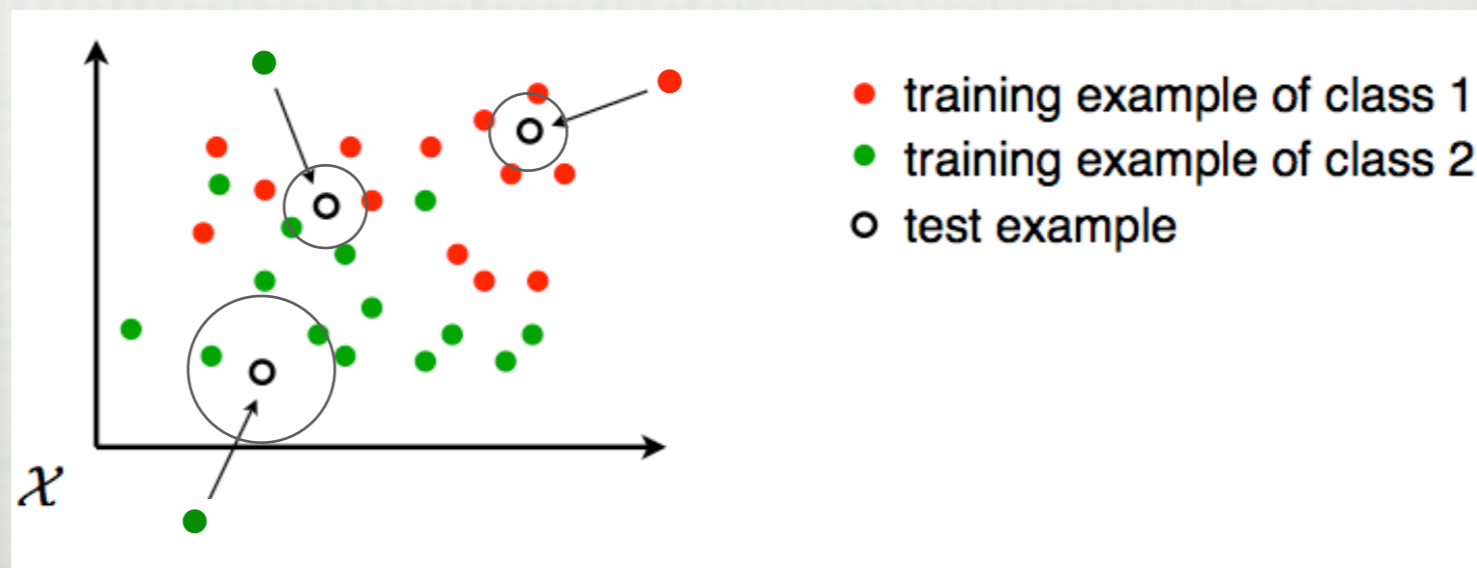
) 1 7 9 1 1 8 5 7 5 0 6 6 0 4 1 2 3 4 4

VIRHEITÄ: 7.6%

(ENSIMMÄISET 5000 MERKKIÄ OPETUSDATAA, SEURAAVAT 1000 MERKKIÄ TESTIDATAA)

K-LÄHIMMÄN NAAPURIN LUOKITIN

- * TOIMINTA: SAMOIN KUIN LÄHIMMÄN NAAPURIN LUOKITTELIJA, MUTTA LÖYDÄ K LÄHINTÄ OPETUSDATAN ESIMERKKIÄ, JA ARVAA LUOKKA TÄMÄN PERUSTEELLA
- * K LÄHINTÄ 'ÄÄNESTÄVÄT' MITÄ LUOKKAA EHDOTETAAN
- * ESIM $K=3$:



NAIVI-BAYES -LUOKITIN

TUTTU SPAM-FILTTERISTÄ!

OPPIMINEN: (ANNETTUNA ISO ESIMERKKIJOUKKO KUVIA+LUOKKIA)

JOKAISALLE LUOKALLE $i=0, \dots, 9$

JOKAISALLE PIIRTEELLE j (TÄSSÄ: 784 PIKSELIÄ)

**ESTIMOI TODENNÄKÖISYYSJAKAUMA $P(x_j=1|i)$ TÄLLE PIKSELILLE
(TODENNÄKÖISYYS ETTÄ KYSEINEN PIKSELI ON 1,
EHDOLLA, ETTÄ LUOKKA ON i)**

LUOKITTELU: (ANNETTUNA UUSI KUVA)

JOKAISALLE LUOKALLE $i=0, \dots, 9$

ASETA TODENNÄKÖISYYS $P(i)=0.1$

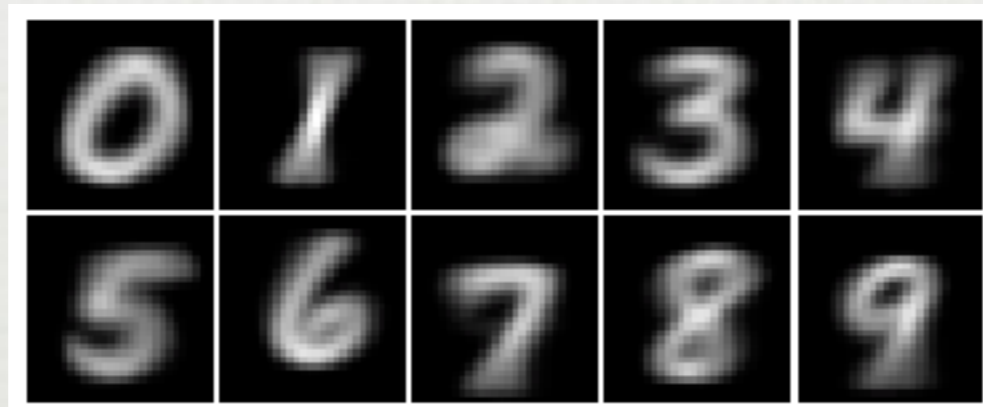
JOKAISALLE PIIRTEELLE j (TÄSSÄ: 784 PIKSELIÄ)

**$P(i) = P(i) * P(x_j=x_j|i)$ (ELI TULO HAVAITTujen PIKSELiEN
TODENNÄKÖISYYKSISTÄ, ANNETTUNA LUOKKA)**

PALAUTA LUOKKA i JOLLA SUURIN $P(i)$.

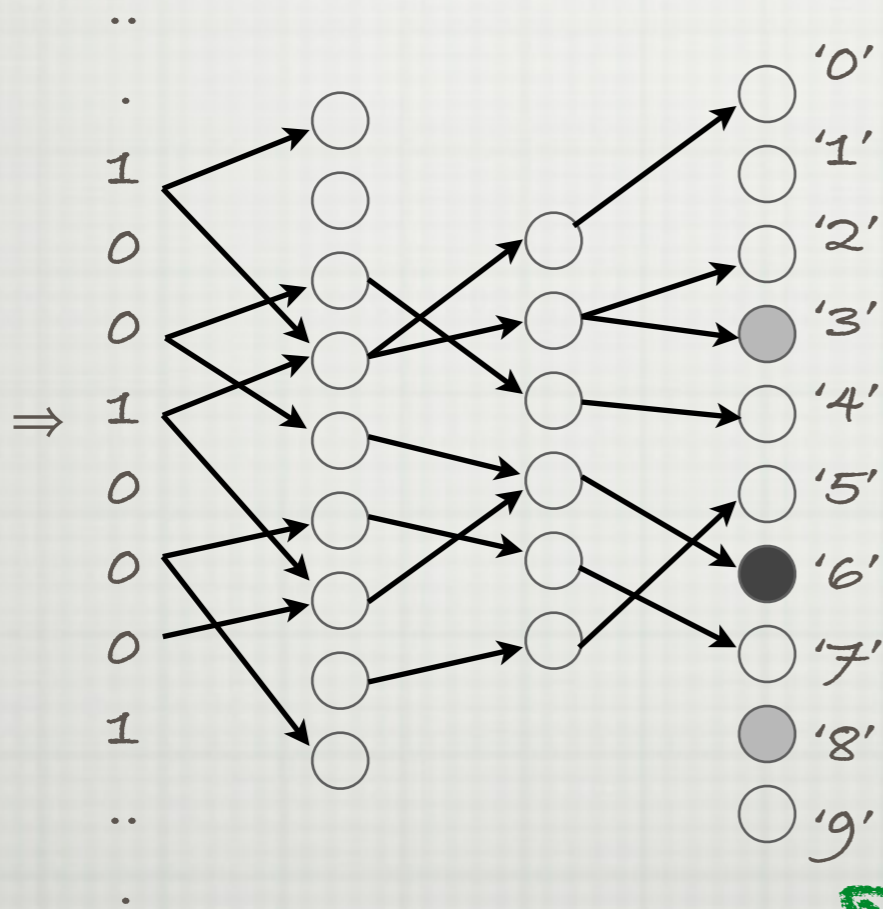
NAIVI-BAYES -LUOKITIN

- * JOKAISEN LUOKAN, JOKAISEN PIKSELIN TODENNÄKÖISYYET VOIDAAN ESITELLÄ YKSINKERTAISESTI NÄIN (HARMAASÄVY = TODENNÄKÖISYYS ETTÄ PIKSELI ON 'PÄÄLLÄ' KYSEISEN LUOKAN KUVILLA)



ELI TÄSSÄ SAADAAN ERÄÄNLAISET 'PROTOTYYPIT' LUOKILLE.
LUOKKA ARVIOIDAAN SEN PERUSTEELLA, KUINKA HYVIN KUVA SOPII NÄIHIN PROTOTYYPPEIHIN.

NEUROVERKKO



JOKAINEN 'SOLU' LASKEE
PAINOTETUN SUMMAN

SYÖTTEISTÄÄN, JA LÄHETTÄÄ
ETEEPÄIN EPÄLINEAARISEN
FUNKTION SIITÄ SUMMASTA

PAINOT OVAT ADAPTIIVISIA, JA
SÄÄDETÄÄN NIIN ETTÄ TULOSTE
ON OIKEA (OHJATTU OPPIMINEN)

TULOSTEKERROS

YHTEENVETO

- * KONEOPPIMISTA KANNATTAÄ KÄYTTÄÄ KUN
 - TEHTÄVÄ ON VAIKEA RATKAISTA 'MANUAALISELLA' OHJELMOINNILLA: OHJELMOIJA EI OSAA ITSE RATKAISTA TEHTÄVÄÄ, TAI EHKÄ OSAA MUTTEI KUITENKAAN PYSTY KERTOMAAN 'MITEN' SE TEHDÄÄN.
 - ESIMERKKEJÄ (DATAA) ON PALJON, JONKA PERUSTEELLA KONE VOI OPPIA ITSE SUORITTAMAAN JOTAIN TEHTÄVÄÄ
 - TARVITAAN ADAPTIIVINEN MENETELMÄ, JOKA MUKAUTUU KÄYTTÄJÄN TOTTUMUKSIIN JA TARPEISIIN
- * ALOITA MÄÄRITTELEMÄLLÄ TEHTÄVÄ, HYVYYSMITTA, JA DATA. TÄMÄNJÄLKEEN VALITSE TEHTÄVÄÄN SOPIVA MENETELMÄ.

MISTÄ LISÄÄ TIETOA?

- * KONEOPPIMINEN, ONGELMAN MÄÄRITTELY, OHJATTU/OHJAAMATON OPPIMINEN:

ESIM: INTRODUCTION TO DATA ANALYSIS USING MACHINE LEARNING BY DAVID TAYLOR (1H 8MIN):

[HTTP://WWW.YOUTUBE.COM/WATCH?V=U4IYSLGNGOY](http://www.youtube.com/watch?v=u4iYSLGNGOY)

- * LÄHIMMÄN NAAPURIN LUOKITTELIJA, K-NN. ESIM:

[HTTP://CSEWEB.UCSD.EDU/~ELKAN/250B/NEARESTN.PDF](http://cseweb.ucsd.edu/~elkan/250B/NEARESTN.PDF)

- * NAIVE BAYES -LUOKITIN. ESIM:

[HTTP://EN.WIKIPEDIA.ORG/WIKI/NAIVE_BAYES_CLASSIFIER](http://en.wikipedia.org/wiki/Naive_Bayes_classifier)

KURSSIT

- * 'INTRODUCTION TO MACHINE LEARNING' (ENGL.)
 - PERIODI II
 - 'ALGORITMIT JA KONEOPPIMINEN'-LINJAN PAKOLLINEN KURSSI
- * JATKOKURSSIT: 'PROBABILISTIC MODELS', 'ADVANCED COURSE IN MACHINE LEARNING', 'DATA MINING', 'SEMINAR: MACHINE LEARNING IN COMPUTER VISION', ...

MISTÄ LISÄÄ TIETOA?

