



# Tämän luennon sisältö

- 1 **Datasta tietoon -johdantoluento**
  - Miksi tällainen kurssi?
  - Kurssin suorittamiseen liittyviä asioita
  - Kurssin sisältö luvuittain

# Miksi tällainen kurssi?

- On haluttu tarjota ohjelman yhteisissä opinnoissa (O-moduuli) kurssi, joka toimii johdatuksena *informaatiotekniikan* opinnoille: tietojenkäsittelytieteen jatkomoduulille A2 ja syventäville A3-moduuleille informaatiotekniikassa *kognitiivisessä neurotieteessä* ja *kieliteknologiassa*.

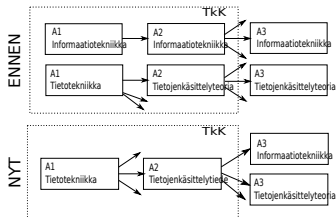
Jatko-moduuli A2 20 op	Perus-moduuli B1 20 op	Kandidaatintyö ja seminaari K 10 op
Perusopinnot P 80 op		Vapaasti valittavat opinnot V 10 op
		Perus-moduuli A1 20 op
		Ohjelman yhteiset opinnot O 20 op

Vapaasti valittavat opinnot W 20 op	Tieteelliset menetelmät M 10 op	Diplomityö D 30 op
Syventävä moduuli A3 20 op	Jatko-moduuli B2 20 op	Erikois-moduuli C 20 op

Kuva: TkT- (180 op) ja DI-tutkinto (120 op)

## Miksi tällainen kurssi? (2)

- HUOM! Informaatiotekniikan A1- ja A2-moduulit ovat poistuneet ja lukuvuodesta 2010-2011 lähtien tarjolla on *tietojenkäsittelytieteen jatkomoduuli A2*, jonka voi valita *tietotekniikan perusmoduulin A1* jatkoksi.



Kuva: (a) Moduulipolku lv 2009-2010 asti; (b) lv 2010-2011 lähtien.

## Miksi tällainen kurssi? (3)

- Bioinformatiikan tutkinto-ohjelmassa kurssi kuuluu laskennallisen ja kognitiivisen biotieteen perusmoduuliin
- Tietojenkäsittelytieteen (TKT) laitos (v. 2008 asti entiset Informaatiotekniikan ja Tietojenkäsittelyteorian laboratorio) on hyvin *tutkimusintensiivinen*. Aalto-yliopiston Perustieteiden korkeakoulussa toimii viisi valtakunnallisista tutkimuksen huippuyksiköistä, joista kaksi TKT-laitoksella:
  1. “Adaptiivisen informatiikan tutkimusyksikkö” (Erkki Oja);
  2. “Algoritmisen data-analyysin huippuyksikkö” (Heikki Mannila)
- Haluamme antaa jo perusopiskelijoille tilaisuuden tutustua meillä tehtävään tutkimukseen

## Miksi tällainen kurssi? (4)

- Ne opiskelijat jotka *eivät valitse* tietojenkäsittelytiedettä tai informaatiotekniikkaa pää- tai sivuaineekseen saavat kuitenkin jonkinlaisen käsityksen siitä mitä ala pitää sisällään
- Ne opiskelijat jotka *valitsevat* tietojenkäsittelytieteen (informaatiotekniikan) pää- tai sivuaineen saavat Datasta Tietoon -kurssissa katsauksen koko kenttään, joka sitten syvenee laitoksen muiden kurssien avulla ja esim. osallistumalla *kesäteekkarina* kesäteekkarina tutkimushankkeisiin (**haku tammikuussa!**)

## Miksi tällainen kurssi? (5)

- Muut aiheeseen liittyvät kurssit (those with *English names* will be given totally in English):

### **T-61.2020 Datasta tietoon harjoitustyö (23.1.2012)**

T-61.3015 Digitaalinen signaalinkäsittely ja suodatus

T-61.3025 Hahmontunnistuksen perusteet

T-61.3040 *Statistical signal modeling*

T-61.3050 *Machine learning: basic principles*

T-61.5010 *Information visualization*

T-61.5020 *Statistical natural language processing*

T-61.5050 *High-throughput bioinformatics*

T-61.5060 *Algorithmic methods of data mining*

T-61.5070 *Computer vision*

T-61.5080 *Signal processing in neuroinformatics*

T-61.5090 *Image analysis in neuroinformatics*

## Miksi tällainen kurssi? (6)

T-61.5100 *Digital image processing*

T-61.5110 *Modeling biological networks*

T-61.5120 *Computational genomics*

T-61.5130 *Machine learning and neural networks*

T-61.5140 *Machine learning: advanced probabilistic methods*

T-61.5150 *Speech recognition*

T-61.5900 Informaatiotekniikan erikoistyö

T-61.5910 *Research project in computer and information science*

T-61.60x0 *Special course in computer and information science I-VI*

T-61.60x0 *Special course in bioinformatics I-II*

T-61.6090 *Special course in language technology*



# Kotisivu, ilmoittautuminen ja suorittaminen

- Kurssin kotisivu <https://noppa.aalto.fi/noppa/kurssi/T-61.2010/>
- Ilmoittaudu WebOodin kautta <https://oodi.aalto.fi/a/>
- Kurssin suoritus: tentti (bonus pisteitä joulutammikuussa) ja pieni harjoitustyö (joka korvattavissa aktiivisella osallistumisella tietokonelaskareihin)

# Käytäntöä

## Luennot

- Luennot maanantaisin ja torstaisin klo 14-16 salissa T1
- Alkupuolen (luennot 1-8, luvut 1-7) luennoi professori Erkki Oja ▶ <http://users.ics.tkk.fi/oja>
- Kaksi viimeistä luentoa (luennot 9-10, luvut 8-9) luennoi dosentti Kai Puolamäki ▶ <http://users.ics.tkk.fi/kaip/>  
(professori Heikki Mannilan osuus)
- Nopassa luentokalvot (kolme formaattia: luvuittain esitys, luvuittain tiivistelmä 4/A4; “koko kirja”)

# Käytäntöä

## Laskuharjoitukset (“paperi”)

- Laskuharjoitukset (2 h / viikko) maanantaisin ja perjantaisin klo 12-14 salissa T4 (vaihtoehtoiset ryhmät). Alkavat pe 4.11.2011. Laskuharjoitusten assistenttina toimivat TkK Maria Osmala ja DI Janne Toivola
- Viisi kierrosta. Sisältävät sekä demotyyppisiä tehtäviä että yhden bonuspistetehtävän / kierros
- Harjoitustehtävät suomeksi ja englanniksi sekä ratkaisut englanniksi löytyvät kurssin kotisivulta

# Käytäntöä

## Laskuharjoitusten bonuspistetehtävät

- Bonuspistetehtävän vastaukset käsin kirjoitettuna A4-kokoiselle paperille. Sivun oikeaan yläkulmaan ISOLLA OPISKELIJANUMERO ja kierros P1, P2, P3, P4 tai P5. Sivun yläosaan myös nimi
- Palautus metalliseen palautuslaatikkoon, T-talon pääaulassa, 3. kerroksessa heti rappusten jälkeen
- Ratkaisuun aikaa noin reilu viikko, DL merkattu tehtäväpapereihin
- Tarkistettujen tehtävien haku laskareista tai 3. kerroksen “Informaatiotekniikan” käytävän lötteröstä

# Käytäntöä

## Pakollinen harjoitustyö

- Pieni Matlab-harjoitustyö (arvioitu työmäärä 8 h) on kiinteä osa kurssia ja sen voi tehdä itsenäisesti 15.1.2012 mennessä.
- Työohje on Nopassa. Aiheena ominaiskasvot (eigenfaces) liittyen pääkomponenttianalyysiin
- Harjoitustyön **voi korvata** aktiivisella osallistumisella ohjattuihin tietokoneharjoituksiin, joita on 2 h viikossa viidesti.

# Käytäntöä

## Tietokoneharjoitukset

- Tietokoneharjoituksia samalla sisällöllä ma 16-18 Maari-A, ti 12-14 Maari-C, ti 14-16 Maari-A, ke 14-16 Maari-A
- Alkavat ma 7.11.2011. Ohjaajina TkK Elina Karp ja TkK Eric Malmi
- Viisi kierrosta. Demoja ja omaa laskentaa tietokoneella. Ei pisteitä, mutta korvaavat pakollisen harjoitustyön
- Tehtävät suomeksi kurssin Noppa-sivulta “viikkoharjoitukset”

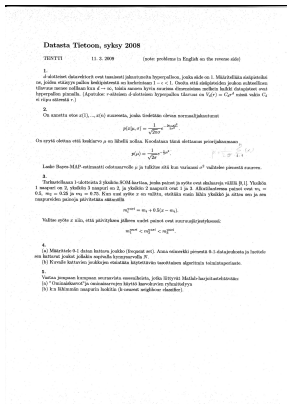
# Käytäntöä

## Tentti

- Ensimmäinen tentti la 17.12.2011 klo 10-13, toinen ke 11.1.2012 klo 13-16 (sali T1) ja viimeinen ennen kesää ke 7.3.2012 13-16 (sali T1).
- Tenttivaatimukset ja luentomateriaali kotisivuilla. Tentissä ei kaavakokoelmaa. Funktiolaskin sallittu.
- Kurssimateriaalina ovat luentokalvot ja harjoitustehtävät ratkaisuiineen, jotka siis saatavana Nopasta
- Viisi tehtävää a 6 p, yhteensä 30 pistettä

# Käytäntöä (2)

## Tentti



Kuva: Tyypillinen tenttipaperi. Lähde: [tentiarkisto.fi](https://tentiarkisto.fi).



# Käytäntöä

## Tentin pisteet ja bonuspisteet sekä arvosanarajat

- Paperilaskareissa 1 bonuspistetehtävä / kierros.  
Bonuspistetehtävät arvostellaan 0 / 0,5 / 1 bonuspiste.
- Lisäksi kurssipalautteen antamisesta saa 1 bonuspisteen.
- Maksimimäärä on 6 bonuspistettä, joka skaalataan tenttipisteiksi jakamalla kahdella (max 3 tenttipistettä).  
Tämä lisätään tenttipisteisiin sillä ehdolla, että tentistä on saanut vähintään 15 pistettä.
- Bonuspisteet ovat voimassa vain joului- ja tammikuun tentissä!
- Arvosanarajat: 0 = 0–14,5p; 1 = 15,0–17,5p; 2 = 18,0–20,5p; 3 = 21,0–24,5p; 4 = 25,0–27,5p; 5 = 28,0–33,0p.

# Käytäntöä

## Tiedotus

- Erityisistä muutoksista ilmoitetaan Nopassa
- Ongelmatapauksissa ota yhteyttä laskuharjoitusassistenttiin
- Kurssin opetushenkilökunnan yhteinen sposti  
`t612010@ics.tkk.fi`

# Kurssin suorittamisesta

- Kurssin osaamistavoitteet:

▶ <https://noppa.aalto.fi/noppa/kurssi/T-61.2010/esite>

- Kurssin jälkeen osaat selittää, kuinka luonnollinen data kuten kuvat, puhe, mittaussarjat esitetään digitaalisesti tietokannassa. Osaat soveltaa tilastomatemattisia ja algoritmisia perusmenetelmiä (yksinkertaisimmassa muodossaan) tällaisen datan käsittelyyn. Kurssin lopussa osaat keskustella asiaankuuluvalla terminologialla, miten "datasta tietoon" käytännössä toteutuu.
- *Mitä "opetustekoja" tarjoamme "oppimistekojanne" varten? Miten tavoitteiden saavuttamista mitataan?*
- **Luennot** tarjoavat pohjan.
- Luentokalvot ovat saatavilla Nopassa. Tutustu aiheeseen etukäteen. Kertaa asioita luennon lopuksi. Kysy epäselviksi jääneistä asioista.

## Kurssin suorittamisesta (2)

- Luentokalvot itsessään eivät ole täydellinen oppikirja!
- **Paperilaskareissa** käydään läpi matemaattista laskemista.
- Laskareiden ymmärtäminen vaatii omaa työtä: tehtävien läpikäymistä itse, välivaiheiden ratkaisua, luennoilla saatua laajempaa ymmärrystä.
- **Tietokonelaskareissa** sovelletaan ja kokeillaan.
- **Tentissä** on viisi kysymystä, joista yksi on suoraan paperilaskareista, ainakin yksi liittyy kurssin loppupuolen lukuihin (Mannila) ja yksi liittyy tietokonelaskareissa sovellettaviin asioihin.

# Kurssipalautteesta

- Kurssin kurssipalaute kerätään joulu-tammikuussa
- Palautteen antamisesta saa **yhden bonuspisteen** joulu- ja tammikuun tenttiin
- Luemme palautteen sekä kurssin päättyessä että uuden alkaessa
- Palautteessa opiskelijoiden suurin yksittäinen haaste: kurssin matemaattisuus → tänä vuonna viikoittaiset bonuspistetehtävät
- Viime syksynä palautteessa kiiteltä mm. laskutupaa, jossa saattoi harjoitella tenttiviikolla (tämä syksy?)

# Kurssin sisältö luvuittain

- 1 Johdanto (1 h)
  - Miksi tällainen kurssi?
  - Käytäntöä
  - Kurssin sisältö
  - Kurssimateriaali
- 2 Datasta tietoon: mitä dataa, mitä tietoa? (1 h)
  - Data-analyysin ongelma
  - Mallit ja oppiminen
  - Esimerkkejä
  - Case study: WEBSOM
- 3 Data vektorina (2 h)
  - Vektorit, matriisit, etäisyysmitat
  - Datan piirreirrotus ja vektorointi
  - Dimensionaalisuuden kirous
  - Esimerkki piirreirrotuksesta: PicSOM

# Kurssin sisältö luvuittain (2)

- 4 Vektoridatan tiivistäminen ja dekorrelointi (2 h)
  - Pääkomponenttianalyysi
  - Esimerkkejä
  - DSS-menetelmä: halutunlaisten aikakomponenttien etsiminen
- 5 Estimointiteorian perusteita (4 h)
  - Perusjakaumat 1-ulotteisina
  - Yleistys vektoridatalle,  $d$ :n muuttujan normaalijakauma
  - Suurimman uskottavuuden periaate
  - Bayes-estimointi
  - Regressiosovitus
  - Esimerkki regressiosta: neuroverkko
  - Esimerkkejä
- 6 Hahmontunnistuksen perusteita (4 h)
  - Johdanto
  - Hahmoalueet, erotinfunktio

# Kurssin sisältö luvuittain (3)

- Lähimmän naapurin luokitin (kNN)
  - Bayes-optimaalinen luokitin
  - Ryhmittelyanalyysi (*c-means*-algoritmi ja hierarkkinen ryhmittely)
- 7 Itseorganisoiva kartta (2 h)
- Perusidea
  - Yhteys biologiaan
  - Suppenevuus 1-ulotteisessa tapauksessa
  - Käytännön valintoja
  - Mihin SOM:ia käytetään?
  - Esimerkkejä
- 8 Hahmojen etsintä diskreetistä datasta (2 h)
- Miten muodostetaan hyviä paikallisia kuvauksia datan osista
  - Algoritmi: Tasottainen algoritmi kattavien joukkojen etsintään



# Kurssin sisältö luvuittain (4)

- 9 Web-etsintämenetelmien algoritmit (2 h)
  - Perusongelmat
  - Linkkirakenteen ottaminen huomioon relevanttien sivujen etsimisessä
  - Algoritmit: Keskukset ja auktoriteetit sekä PageRank (Google)