is this a or b? (logistic/classification)

is this weird? (anomoli)

how much/many? (regression)

how is this organized? (cluster)

What do i do next? (reinforcement)

underfitting

overfitting

bias/variance

visualiseringar

- outputs multiclass ROC

receiver operating characteristic

- output -confusion matrix output

linear regression - best fit.

y = bx + a

b = formel

a = y.vekt - b \* xvekt.

best fit = minsta kvadratiskt avstånd mellan exempel och linje.

confidensintervall = inom 95% inom 100-120.

logistic

P = e(a+bX) / 1 + e(a+bX)

accuracy

precision

CNTK

networks = functions

tensor = karta av matriser

reader, network, trainer

nodes = functions

defferd computations

h1 = sigmoid(W1x + b1)

h2 = sigmoid(w2H1+b2)

P =

cross\_entropy

learner ? sgd(P.

CNN

- pooling (max-pooling)

- convulition (mult med transformerande matris)

data

relevant

connected (inte massa blanka kolumner, många helt kompletta rader)

accurate

enough

bilder hundratusentals eller miljoner

bilder från alla håll på objektet

transferlearning, klara sig med betydligt färre bilder för att skilja på två objekt.

3 alternativ känna igen lövträd

- custom visions

- neuralt nätverk

- transfer learning med neuralt nätverk. Göras i R eller Phyton.

-edge detections

- inspect, histogram och andra visualisering.

- datawrangling kan jag använda beskrivningsfältet och få ut bra features.

Hantera dubletter

text clustering - alla varianter av en benämning.

Hantera json.

Blir kod. Generate data access code-file. Går inte just nu att exportera pythonkod, finns bakom.

\* luis.ai. - intention - hitta station t ex. språkförståelse. officellt finns inte svenska än, men funkar ofta ändå. utveckla i c# eller nodejs. En svensk bot pågång - logic apps - söker i andra system.

definiera fråga svar. bitlocker recovery key. letar frågor och matchar till svar. definieras frågor och svar manuellt. finns som cloud.

supervised - finns facit

70% av tiden data preprocessering

features/targetvalue

features/predictor

1 välj features

2 splitta träning och valideringsdata

förbättringar:

- fler featurs, skapa features, mer data, byta algoritm, konfigurera hyperparametrar

hantera collinearity, autocorreltation, heteroskedasticity. finns detta behöver det hanteras, modeller hanterar inte detta.

azure ml services (inklusive workbench)

finns som preview sedan 2 v.

regression

classification

clustering

börja ofta med linear och logistic regression

Data science VM

sätta igång när man behöver - annars kan det bli kostsamt.

har massor av biblioteket, visualstudio, power bi, (tensorflow?)

finns för windows, linux etc.

cognitive services

\* bildanalys - se olämpligt innehåll, se känsla, känna igen person, indexera vido, can customseras,

\* tal - känna igen ord, ej svenska så bra än,

\* language - stavning, språkanalys, hitta nyckelord, översätt, get sentiment, tokenize, identiferia ämnen finns även on prem - ml server och sql seriver.

\* kunnande - vetande

\* search - bing, bing bild etc,

\* custom vision - ladda upp bilder, två kategorier, tryck träna... tar 5 minuter. Retailer - känna igen bilar.

- ladda upp bilder - tagga - träna. objektklassifcering, inte objektdetection. Använder transerlearning. Deep laerning modell som är förtränad. ml server (sql server) finns där också, måste dock göras med r och python, snarare än ett gui.

ecosystem ml

data i någon form

moln/prem

tränna modell i moln, men kör modell on prem

edge - i bil eller device.

fast and agile

innbyggd ai

api - cognitive services. färdiga api:er (som inte kräver kunskap i datascience)

algoritmer (ml framwork - tensorflow, CNTK, scikot-learn, Ml algoritms - revoscalr, revoscalepymsml, dmtk)

tools ( visual studio code, azure machine learning - algoritmer och ramverk)

datalayer (sql server, data science vm, visual studio code, azure data lake, hdinsight, ml server)

azure infrastruktur (GPU, FPGA)

lättast

- power bi färdiga fuktioner

- cognitive services färdiga funktioner

- azure machine learning studio

- azure machine learning workbench med ramverk

svårast

machine learning studio och azure machine learning, två olika ingångar, studio ger mycket stöd, lättare att komma igång med.

väldigt stora datamängder - azure data lake (store och analytics), hdinsight (hadoop, spark, även inbyggda ml verktyg - R, spark ml, cluster eller

kör ml studio för nybörjaren - workbench där sker utvecklingen, skalar. workbench har wrangling funktioner.

Azure hdinsights?

MMLSpark abstraherar bort mycket mycket som annars måste kodas förhand (splitta, göra om till kategoriska)

RNN

t1, t2, t3, t4

är input

Process mining

heuristisk process mining

processlearning.org

det går att kombinera trad. teknik (t ex open cv för bilder) och deep learning.

cross entropy

log likelyhood -

\* botframework - botbuilder. använder luis. bot connect till skype, bing,

idag -

har exjobb pågång

vilken kurs?

andrew ng

data science essentials

LSTM

tidsserier

som tal, när och långtidsminne.

bilder

-objektdetektering - var i bilden?

djuphet

resnet - 150 dolda lager

verktyg prestanda

caffe, cntk, tensorflow, torch

cntk - snabbast

azure batch kommer

skapa en beräkningsinstans i azure, som man lägger upp en instalns av sql server i azure.

går att köra tensorflow och cntk med gpuer lokalt.

Säljprocess

Ställa upp en tes

machine learning studio

används av 1000 tusentals modeller

men kunder vill: bättre kontroll

därför azure machine learning services = workbench, kunna köra modeller där man vill.

1 experimentation. cmd, ide, workbench, vs code - köra modeller - local, dvsm, etc, finns alla ramverk

2 modell management, skicka via docker, vidare till azure eller vad man vill. hanteras med cmd.

azure machine learning - tvätta data, träna modell, utvärdera, deploya modell.

datawrangling, ai-driven. Slipper regex, strängmanipulering.

att bara släppa lös ai på en text, kan imponera vid demo, men är kommer inte fungera tillräckligt bra praktiskt i dagligt arbete.

azure search

indexerar och gör sökningar möjliga

skicka en stor knowledge base.

jönsson

kan hjälpa med kunder

targets

local

datascience vm

docker