

Tehtävä 1. Koneoppimista (2 pistettä)

1. (1 piste) Perusasetuksilla monikerrosperspertroni antoi virheeksi 47.06 %. Lisäämällä opetuskelten (`trainingTime`) määrää 500:sta 5000:en, virhe vain kasvaa 48.53 prosenttiin. Asettamalla asetuksen `hiddenLayers` arvoksi esimerkiksi '1', eli valitsemalla yksi piilokerros, jossa on vain yksi neuroni, saadaan virheeksi 45.59 %. Kahdella piiloneuronilla saadaan virhe 48.53, joka on ilmeisesti oletusasetus, koska tämä saatiin jo aiemmin. Kolmella piiloneuronilla virhe on 30.88 % ja neljällä 5.88 %.
2. (1 piste) Lähimmän naapurin luokitin (`Lazy->IB1`) toimii hurjan hyvin: luokitustarkkuus on 100 %. Lähimmän naapurin luokitin tuntuukin lähtökohtaisesti hyvältä idealta, koska esimerkit ovat melko tiheässä ja vierekkäisillä esimerkeillä on sama luokka-arvo.

Päätöspuu (`Trees->J48`) oletusasetuksilla tuottaa virheen 11.77 %. Päätöspuu jakaa x, y -avaruuden rekursiivisesti akselien suuntaisilla suorilla, esim. $x > 1.5$ ja tekee luokituspäätöksen riippuen siitä, kummalla puolella suoraa esimerkit ovat. Tämä ei tunnu hyvältä keinolta, kun ottaa huomioon, että kahta spiraalia (eri luokkia) ei voi erottaa toisistaan akselien suuntaisilla suorilla ilman, että käyttää erittäin montaa suoraa. Saatu luokitustarkkuus on sikäli ehkä yllättävänkin hyvä. (Tästä lisää *Introduction to Machine Learning* -kurssilla.)

Tehtävä 2. Hahmontunnistus (2 pistettä)

- a) (1 piste). Moni oli saanut erittäin hyviä tuloksia esineitä tai rakennuksia esittävistä kuvista, jotka on otettu enemmän tai vähemmän samasta kuvakulmasta. Ihmisistä tulokset olivat merkittävästi huonompia.
- b) (1 piste). SURF ei näytä pahemmin häiriintyvän rotaatioista, rajauksista, skaalauksista tai lievästä sumentamisesta tai värien muokkaamisesta.

Tehtävä 3. Luonnollisen kielen käsittely (2 pistettä)

a) (1p) CYK-algoritmissa laskettava taulukko:

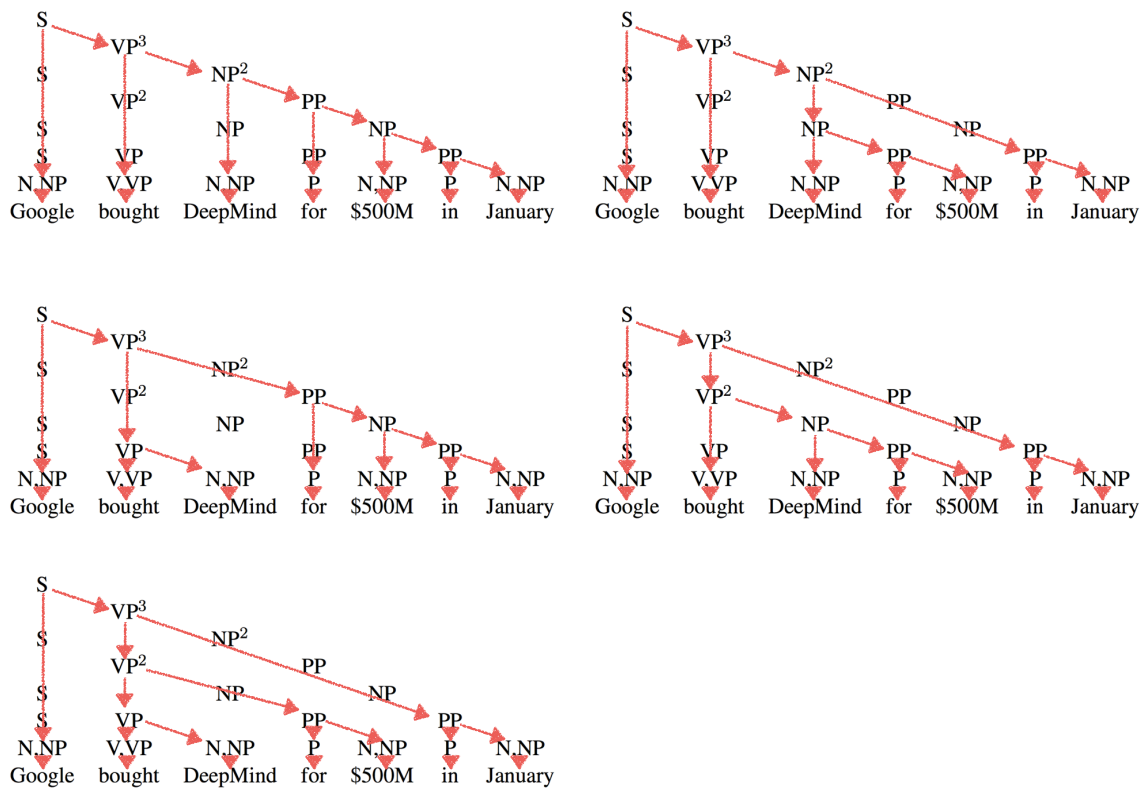
S						
	VP ³					
S		NP ²				
	VP ²		PP			
S		NP		NP		
S	VP		PP		PP	
N,NP	V,VP	N,NP	P	N,NP	P	N,NP
Google	bought	DeepMind	for	\$500M	in	January

Merkinnät muotoa VP³ tarkoittavat, että ko. välike voidaan muodostaa useammalla kuin yhdellä tavalla, mikä johtaa useaan vaihtoehtoiseen tapaan muodostaa jäsennyspuu. Esimerkiksi solun (2, 7) välike VP saadaan kolmella eri tavalla: (i) soluista (2, 5) (VP) ja (6, 7) (PP) säännöllä $VP \rightarrow VP PP$; (ii) soluista (2, 3) (VP) ja (4, 7) (PP) säännöllä $VP \rightarrow VP PP$; (iii) soluista (2, 2) (V) ja (3, 7) (NP) säännöllä $VP \rightarrow V NP$.

Kolmesta vaihtoehtoisesta tavasta muodostaa solun (2, 7) välike VP sekä vaihtoehto (i) että vaihtoehto (iii) liittyvät edelleen välikkeisiin, jotka voidaan muodostaa kahdella eri tavalla. (Solun (2, 5) välike VP saadaan joko soluista (2, 3) (VP) ja (4, 5) (PP) tai soluista (2, 2) (V) ja (3, 5) (NP).)

Jäsennyspuuta tulee edellä kuvatuista vaihtoehtoista yhteensä viisi, ks. kuva 1. Luontevin jäsennyspuu (alimmainen) vastaa tulkintaa, että Google teki jotain, missä tekeminen oli ostamista ja ostamisen kohteena oli DeepMind ja lisämääreinä \$500M hinta sekä ajan-kohta tammikuussa. Sen sijaan ylärivin vasemmanpuoleinen ja keskirivin vasemmanpuoleinen jäsennyspuu sisältävät alipuun (*\$500M (in January)*), joka antaa ymmärtää, että tammikuu liittyy 500 miljoonaan dollariin. Oikea tulkinta on, että tammikuu liittyy ostamiseen. Oikeanpuoleiset puut puolestaan sisältävät alipuun (*DeepMind (for \$500M)*), joka tulkittaisiin siten, että \$500M on DeepMindin ominaisuus, joka olisi riippumaton ostamisesta.

b) (1p) Esimerkkiratkaisu tulee kurssin sivulle.



Kuva 1: Lauseen "Google bought DeepMind for \$500M in January" mahdolliset jäsennyspuut. Niistä alimmainen on semanttisesti luotevin.