

Dokumentazioa Proiektua

Ieltzu Irazu, Mikel de Velasco, Jorge Nieto

29 de enero de 2010



1 Azaleko Orria / Aurkibidea

1. AZALEKO ORRIA / AURKIBIDEA	2
2.	

2 Esparru Teorikoa

1. Dokumentazio eta sintesi lana landutako ikasketa teknikari buruz, alegia:

- (a) Support Vector Machines(SVM)
- (b) **Multilayer Perceptron (MP)**
- (c) Bayes Network (BayesNet)

Gure kasuan **Multilayer Perceptron** estimatzailea ikastea tokatu zaigu. **Multilayer Perceptron** ez da estimatzaile simple bat eta horregatik **OneR** estimatzailearekin konparatuko dugu (estimatzaile simple dena).

2.1 Multilayer Perceptron

‘Multilayer Perceptron’ ingurune aldaketen aurrean aldatzen duen sare neuronal artifiziala da, eta sarrerako datu batzuetatik irteera egoki batzuetara aldatzen du. ‘Multilayer Perceptron’ sarea hiru geruza mota desberdinetan banatzen da:

- (a) **Sarrera geruza:** Sarean sarrera patroiak sartzen duten neurona multzoa. Neurona hauek ez dute prozesamendurik.
- (b) **Geruza ezkutuan:** Erdibideko neuronak dira, non bakoitzaren sarrerak aurreko geruza batetik datoz eta irteerak hurrengo geruza batera doaz.
- (c) **Irteera geruza:** Zeinen irteera-balioak sare guztiaren irteerekin bat datozen neuronak.

‘Multilayer Perceptron’ atzerako hedapen ikasketa entrenatzeko teknika erabiltzen duen sarea da. Metodo honek ‘Estandar Linear Perceptron’ metodoaren aldaketa bat da eta linelaki banatzeko modukoak ez diren datuak bereiz ditzake.

ATZERAKO HEDAPENEAREKIN IKASTEN:

Ikasketa datuak prozesatu eta gero haien pisua aldatzen denean gertatzen da, espero zen errore kopurua egon den errore kopuruarekin konparatuz.

Algoritmoa: Errorea Kalkulatu
(Nodo Kopurua: N | Data Puntua: j | TargetValue: d | Perceptroiak emandako balioa: y)

Aurre-baldintzak: $n \in N$

Eraitza: $e_n(j)$

Begin :
$$e_n(j) = d_n(j) - y_n(j)$$

End

Errorea irteeratik sarrera joango balitz bezala ulertu daiteke, horregatik hartzen du backpropagation izena.

PERCEPTROIA ENTRENATU:

Perceptroiek hiperplano bat definitzen dute, eta neurona sarea perceptroiaaren hiperplano hori implementatzen dute. Datu batzuk sarreratzat erabiliz, pisu balioak kalkulatu egin daitezke konexiorik gabe eta konektatzean

pertzeptroiak irteera balioak kalkulatu ditu.

Pertzeptroiari ez zaio lagin osoa pasatzen lehenengo aldian, honen zati bat ematen zaio. Sareak bere aldagaiak kasu bakoitzaren ostean eguneratzea da helburua, sarea gutxika definituz eta hobetuz. Errorea ez da lagin osoaren gainean kalkulatzeko, baizik eta jasotako instantzia bakoitzeko errore hori txikitzen ahalegintzen da.

2. Diseinua laburbildu eta atazen banaketa taldekideen artean eman. Javan garatu den programaren diseinua eta inplementazioari buruzko xehetasun aipagarriak.

3 Esparru Esperimentala

1. Datuak: datuen deskribapen kualitatibo eta kuantitatiboa 2. ariketako emaitzak sartu.

	Pre-processed Data		
	Train	Dev	Test
Nominal attributes	2	2	2
Numeric attributes	1	1	1
Attributes Total	3	3	3
Instances V1	11326	45	0
Instances V2	11202	17363	0
Instances total	22528	17408	24220

2. Aurre-prozesamendua: erabilitako filtroak aukeratzeko motibazioa azaldu eta 3.ariketako emaitzak eman.

	Class	Hold Out (train vs. dev)		
		Precision	Recall	F-Measure
Baseline	V1	0.3333333333333333	0.3777777777777777	0.3541666666666667
	V2	0.9983868179985	0.99804181305	0.9982142857142858
	W.Avg.	0.8803191444948786		
Model	V1	0.0	0.0	0.0
	V2	0.9974149816176471	1.0	0.9987058180667797
	W.Avg	0.9961241451685142		

3. Emaitza esperimentalak: ereduak eskaintzen duten kalitatea aztertu. ...

	Class	Resubstitution error			10-Fold Cross Validation		
		Precision	Recall	F-Measure	Precision	Recall	F-Measure
Baseline	V1	0.61136 9719539 1598	0.92401 7236830 5338	0.735861610 113107	0.6254901 96078431 4	0.533022 60135432 24	0.575566212 4305588
	V2	0.96202 1978021 978	0.76618 2391037 9836	0.853005943 6811849	0.8244395 95318389 2	0.872956 41519341 85	0.848004624 9851217
	W.Avg	0.8196513709397332			0.7704330807854539		
Model	V1	0.0	0.0	0.0	0.0	0.0	0.0
	V2	0.71526 9431089 7436	1.0	0.834002423 3222872	0.7152694 31089743	1.0	0.834002423 3222872
	W.Avg	0.5965364388571999			0.5965364388571999		

4. Exekutagarrien exekuzio adibide bat (ikusi 5.1. atala).

4 Ondorioak eta etorkizunerako lana

Lan honetan Multilayer Perceptron-ren parametro asko ekortu ditugu eta beraz oso modelo sendoa lortu dugu. Nahiz eta parametro asko ekortu izan, konturatu gara ez dela asko hobetzen modeloa. Eskatzen zeneko ia-ia guztia egin dugu eta beraz ez ditugu asko egiteko, baina etorkizuneko lan bezala beste Baseline algoritmoak ekortzea eta konparaketa berriak egitea izango zen.

5 Bibliografia

Bibliografia edukia indartu edo sostengatzeko emanda dago. Ezinbestekoa da iturri bibliografikoak aipatzea erabili edo horietara jotzen garen puntuan bertan. Ez ahaztu irudien iturria aipatzen. Testuan zehar esplizituki aipatu ez den iturririk ez sartu Bibliografia atalean. Iturria aipatzean zehaztu ahalbait gehien (liburuko kapitulua edo atala eman), honela beste irakurle batek sakondu ahal izango du iturri horiek kontsultatu.

- **MULTILAYER PERCEPTRON:** Data Mining - Practical Machine Learning Tools and Techniques (3rd Ed) (Page 232)
- **MULTILAYER PERCEPTRON:** en.wikipedia.org/wiki/Multilayer_perceptron
- **MULTILAYER:** en.wikipedia.org/wiki/Perceptron

6 Balorazio Subjektiboa

- Ieltzu: Nire ustez Atazarekin lortu behar genueneko helburuak lortu dira. Denbora asko eman dut bereziki txostena egiten, gauza asko azaldu behar direlako. Ikasketarako denbora nahikotzo ere eman dut. Multilayer Perceptron nola doan jakitea denbora asko eraman dit. Diseinua egiteko ez da arazo larririk egon. Diseinua egiterakoan beharbada modeloa egiteko atala zailena izan da, luzera handienekoa da ere. Taldelana oso garrantzitsua izan da. Beti bezala, gure taldean oso ondo moldatzen gara hirurok batera eta pozik nago egindako lanarekin. Interes handiena beharbada lan handia izatearena izango da, ez genuelako hain handia den bestelako lanik egin arlo honetan.
- Jorge:

- Mikel:

(borondatezkoa) eranskin batean atazari buruzko hausnarketa egin. Honako puntuak lagungarriak izan daitezke baina ez dira derrigorrezkoak, beste batzuk ere sar daitezke.

1. Atazarekin lortu nahi ziren helburuak lortu dituzue? (ikusi 1. atala)
2. Batazbestean zenbat denbora eman duzue atazea lanean? Desglosatu: ikasketarako denbora, bilaketa bibliografikoa, softwarearen diseinua eta inplementazioa, txostena.
3. Talde-lana erabilgarria izan da ataza ebazteko?
4. Zer sortarazi dizue interesik handiena? Ikasleengan interesa eta motibazioa pizteko iradokizunak eman.

7 PROBANDO

Algoritmoa: NombreAlgoritmo (Entrenatzeko Instantziak: Z)

Aurre-baldintzak: Z $|Z| = N$

Eraitza: Erabaki Zuhaitza: T

Begin :

```
if(  $Z$  multzoko instantzia guztiak klase berekoak dira ( $C$ ) ){  
    .   return  $T$ (nodo terminala (hostoa)  $C$ -rekin etiketatua)  
} else {  
    .   Haukeratu informazio gehien eskaintzen duen aldagaia  $Xr(Xr =$   
    { $xr^1, \dots, xr^{nr}$ )  
    .   ....  
    .   return  $T$ (lotu error  $Xrnr$  azpi-zuhaitzarekin:  $T1, \dots, Tnr$ )  
}
```

End