

ΑΝΑΦΟΡΑ ΓΙΑ ΤΗΝ 1Η ΕΡΓΑΣΙΑ ΣΤΗΝ ΜΗΧΑΝΙΚΗ

ΜΑΘΗΣΗ

ΚΑΘΗΓΗΤΗΣ	ΚΩΝΣΤΑΝΤΙΝΟΣ ΜΠΛΕΚΑΣ	
ΑΜ	4679	4065
ΜΕΛΗ ΟΜΑΔΑΣ	ΚΑΖΑΚΙΔΗΣ ΘΕΟΧΑΡΗΣ	ΚΑΖΑΚΙΔΗΣ ΚΩΝΣΤΑΝΤΙΝΟΣ
ΕΤΟΣ	3 ^ο	5 ^ο

Για την υλοποίηση της εργασίας χρησιμοποιήσαμε την πλατφόρμα `juryter`. Διευκόλυνε την υλοποίηση καθώς σπάει τον κωδικά σε κελιά και μπορούμε να τρεχουμε ξεχωριστά κομμάτια κωδικά αλλά σε ένα script

Από το `datasheets` ορίσαμε τα δεδομένα (`bone_length,rotting_flesh,hair_length,has_soul,color`) ως x (αυτά τα χαρακτηριστικά που θα κάνουν το `train`, το `id` δεν επηρεάζει το αποτέλεσμα αρα δεν το περνουμε υποψιν) και το `type` ως y που είναι και το `target` μας

Εχουμε τα δεδομένα μας σε κανονικοποιημένη μορφή επομένως μπορούμε να τα χειριστούμε πιο ευκολά και να βγάλουμε καλύτερα αποτελέσματα και δεν χρειάζεται να τα μετατρέψουμε αφού είναι ήδη

Γράψαμε σε μορφή κωδικά τις συναρτήσεις που μας δώσατε στο pdf καθώς μας είναι πολύ χρησιμες στην συνέχεια και στην υλοποίηση του project

Επειδη το project βασίζεται σε μαθηματικά μοντελά πρέπει να μετατρέψουμε τα `string` σε αριθμούς (όπως το `color` και το `type`) για να τα φορτώσουμε στο x, y μοντελό αρα τα αντιστοιχουμε με αριθμούς

****Αμα δεν τρεχει στην εντολη `train_set=pd.read_csv('../input/ghouls-ghost-goblin/train.csv')` παρακαλουμε βαλτε το δικος σας path για το `train.csv`.Επισης χρησιμοποιησα το `train.csv` και όχι το `train.csv.zip` ,μπορει να φαινονται αυτονοητα αλλα οφειλουμε να τα αναφερουμε**

Μεσω τις βιβλιοθηκης `sklearn` και συγκεκριμενα `from sklearn.model_selection import train_test_split` εχουμε μια βοηθητικη συναρτηση που σπάει τα x και y δεδομένα σε `train` και `test` αντιστοιχα

αφου όταν κανουμε εκπαιδευση θελουμε να σπασουμε τα δεδομενα σε δυο ομαδες x_{train}, y_{train} και x_{test}, y_{test} .Επισης βαζουμε το stratify για εξασφαλισουμε μια ισορροπια στο μοντελο μας αρα θα εκπαιδευσουμε εξισου καλα ολα τα ειδη monster

KNN Μεθοδος:

Πρόκειται για ένα σχετικά απλό αλγόριθμο ο οποίος ελέγχει όλες τις περιπτώσεις και ταξινομεί τις καινούριες σύμφωνα με τα γειτονικά στοιχεία. Για να γίνει η ταξινόμηση αυτή χρησιμοποιείται μία συνάρτηση απόστασης

Η 1η μεθοδο που χρησιμοποιουμε είναι η knn όπως γνωριζουμε το knn δεν εχει εκπαιδευση για αυτό χρησιμοποιει την ευκλειδια απόσταση γιατι ουσιαστικα μετραει την αποσταση μεταξυ των features των monsters και αμα η αποσταση είναι μεσα στην γειτνια που εχει σχηματισει τοτε βρισκουμε οντως το κάθε type.Για κάθε τυπου monster υπαρχει διαφορετικη γειτνια.

Δημιουργησαμε έναν πινακα με τις τιμες του k και μια loop που φορτωνει τα δεδομενα στο knn κανει τις πραξεις βρισκει τις συσχετισεις με διαφορετικο k κάθε φορα και στην συνεχεια κανει τις προβλεψεις με τα x_{test} .Τα x_{test} δεδομενα βγαζουνε y_{test} αποτελεσματα αλλα αυτά τα y είναι τα predictions.Για να δουμε αμα οντως πετυχε το αποτελεσμα y κανουμε το πινακα confusion matrix δηλαδη οσα είναι TP(πανω στην διαγωνιο).Ετσι βρισκουμε τα κελια που μας δινουν το TP,FN,TN,FP και το μεταφραζουμε σε κωδικα (ξεχωριστα για ghost,goblins,ghoul) .Επειτα καλουμε τις συναρτησεις που φτιαξαμε παραπανω για το (accuracy,precision,recall,f1) και τελος αφου κανουμε για τοκαθενα ξεχωριστα κανουμε την διαδικασια για το συνολικο και βρισκουμε τα συνολικα αποτελεσματα(ουσιαστικα κανουμε ένα μεσο ορο το αποτελεσματος)

		Predicted Class		
		Positive	Negative	
Actual Class	Positive	True Positive (TP)	False Negative (FN) Type II Error	Sensitivity $\frac{TP}{(TP + FN)}$
	Negative	False Positive (FP) Type I Error	True Negative (TN)	Specificity $\frac{TN}{(TN + FP)}$
		Precision $\frac{TP}{(TP + FP)}$	Negative Predictive Value $\frac{TN}{(TN + FN)}$	Accuracy $\frac{TP + TN}{(TP + TN + FP + FN)}$

Τα αποτελεσματα για KNN:

Παρατηρούμε ότι το καλύτερο accuracy το πετυχαίνουμε με k=5 και συγκεκριμένα είναι Accuracy = 0.8103448275862069

Ghost k: 1 Accuracy: 0.8717948717948718 Precision:0.8888888888888888 Recall: 0.6666666666666666 F1 score: 0.761904761904762	Ghost k: 3 Accuracy: 0.9 Precision:0.7857142857142857 Recall: 0.9166666666666666 F1 score: 0.8461538461538461
Goblin k: 1 Accuracy: 0.6578947368421053 Precision:0.5 Recall: 0.7692307692307693 F1 score: 0.6060606060606061	Goblin k: 3 Accuracy: 0.6578947368421053 Precision:0.5 Recall: 0.38461538461538464 F1 score: 0.4347826086956522
Ghoul k: 1 Accuracy: 0.7894736842105263 Precision:0.7777777777777778 Recall: 0.5384615384615384 F1 score: 0.6363636363636364	Ghoul k: 3 Accuracy: 0.7105263157894737 Precision:0.5714285714285714 Recall: 0.6153846153846154 F1 score: 0.5925925925925927
Total k: 1 Accuracy: 0.7739130434782608 Precision:0.6578947368421053 Recall: 0.6578947368421053 F1 score: 0.6578947368421053	Total k: 3 Accuracy: 0.7586206896551724 Precision:0.631578947368421 Recall: 0.631578947368421 F1 score: 0.631578947368421

Ghost k: 5 Accuracy: 0.95 Precision:0.8571428571428571 Recall: 1.0 F1 score: 0.923076923076923	Ghost k: 10 Accuracy: 0.9047619047619048 Precision:0.75 Recall: 1.0 F1 score: 0.8571428571428571
Goblin k: 5 Accuracy: 0.7105263157894737 Precision:0.5833333333333334 Recall: 0.5384615384615384 F1 score: 0.5599999999999999	Goblin k: 10 Accuracy: 0.6842105263157895 Precision:0.5384615384615384 Recall: 0.5384615384615384 F1 score: 0.5384615384615384
Ghoul k: 5 Accuracy: 0.7631578947368421 Precision:0.6666666666666666 Recall: 0.6153846153846154 F1 score: 0.64	Ghoul k: 10 Accuracy: 0.7894736842105263 Precision:0.7777777777777778 Recall: 0.5384615384615384 F1 score: 0.6363636363636364
Total k: 5 Accuracy: 0.8103448275862069 Precision:0.7105263157894737 Recall: 0.7105263157894737 F1 score: 0.7105263157894737	Total k: 10 Accuracy: 0.7966101694915254 Precision:0.6842105263157895 Recall: 0.6842105263157895 F1 score: 0.6842105263157895

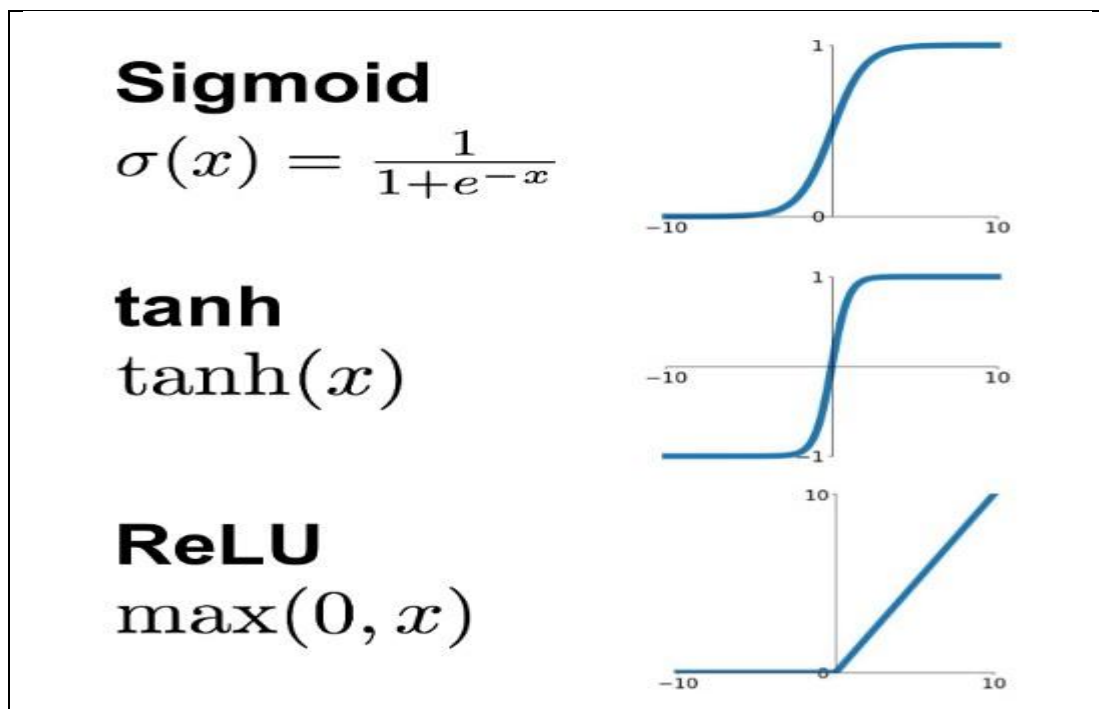
Total k: 1 Accuracy: 0.7739130434782608 Precision:0.6578947368421053 Recall: 0.6578947368421053 F1 score: 0.6578947368421053	Total k: 3 Accuracy: 0.7586206896551724 Precision:0.631578947368421 Recall: 0.631578947368421 F1 score: 0.631578947368421
Total k: 5 Accuracy: 0.8103448275862069 Precision:0.7105263157894737 Recall: 0.7105263157894737 F1 score: 0.7105263157894737	Total k: 10 Accuracy: 0.7966101694915254 Precision:0.6842105263157895 Recall: 0.6842105263157895 F1 score: 0.6842105263157895

Νευρωνικά Δίκτυα Μεθοδος:

Κανουμε τρεις εκπαιδευσεις ένα κρυφο επιπεδο με 50,100 και 200 νευρωνες και τρεις εκπαιδευσεις με δυο κρυφα επιπεδα με (50 νευρωνες το πρωτο και 25 το δευτερο),(100 νευρωνες το πρωτο και 50 το δευτερο),(200 νευρωνες το πρωτο και 100 το δευτερο). Κανουμε μια λουπα που θα κανει δοκιμες αντιστοιχα για κάθε περιπτωση νευρωνων.

Χρησιμοποιούμε για συναρτηση ενεργοποίησης αναμεσα στους νευρωνες,την \tanh^{**} .Χρησιμοποιουμε ως optimizer το Stochastic Gradient Descent (sgd) ώστε να γινονται καλυτερες συσχετισεις μεταξυ των δεδομενων.Επειτα φορτωνουμε τα δεδομενα στο νευρωνικο δίκτυο,κανουμε τις προβλεψεις και μετα κανουμε confusion matrix όπως και στην παραπάνω μεθοδο

**** παρατηρησαμε ότι με την logistic μερικα αποτελεσματα εβγαζαν σφαλαματα όπως nan και ετσι πειραματιστηκαμε και βαλαμε σαν συναρτηση ενεργοποιησης την \tanh και ειδαμε καλυτερα αποτελεσματα,επισης χρησιμοποιησαμε την relu καπεριμεναμε ότι θα βγαλει ακομα καλυτερα αποτελσματα καθως ειδαμε ότι θεωρητικα είναι η πιο αποδοτικη αλλα στην περιπτωση αυτή η \tanh ηταν καλυτερη**



Τα αποτελεσματα των νευρωνικων δικτυων:

1)Με ένα κρυφο επιπεδο

<p>Ghost Hidden layers: (50,) Accuracy: 0.7 Precision:0.5 Recall: 0.3333333333333333 F1 score: 0.4</p> <p>Goblin Hidden layers: (50,) Accuracy: 0.71052631578947 Precision:0.625 Recall: 0.3846153846153846 F1 score: 0.47619047619047</p> <p>Ghoul Hidden layers: (50,) Accuracy: 0.55263157894736 Precision:0.409090909090909 Recall: 0.6923076923076923 F1 score: 0.51428571428571</p> <p>Total Hidden layers: (50,) Accuracy: 0.65517241379310 Precision:0.47368421052631 Recall: 0.473684210526315 F1 score: 0.47368421052631</p>	<p>Ghost Hidden layers: (100,) Accuracy: 0.825 Precision:0.777777777777777 Recall: 0.583333333333333 F1 score: 0.666666666666666</p> <p>Goblin Hidden layers: (100,) Accuracy: 0.63157894736842 Precision:0.42857142857142 Recall: 0.230769230769230 F1 score: 0.3</p> <p>Ghoul Hidden layers: (100,) Accuracy: 0.60526315789473 Precision:0.454545454545454 Recall: 0.769230769230769 F1 score: 0.57142857142857</p> <p>Total Hidden layers: (100,) Accuracy: 0.68965517241379 Precision:0.52631578947368 Recall: 0.526315789473684 F1 score: 0.52631578947368</p>	<p>Ghost Hidden layers: (200,) Accuracy: 0.82926829268292 Precision:0.72727272727272 Recall: 0.666666666666666 F1 score: 0.6956521739130</p> <p>Goblin Hidden layers: (200,) Accuracy: 0.63157894736842 Precision:0.4 Recall: 0.153846153846153 F1 score: 0.22222222222222</p> <p>Ghoul Hidden layers: (200,) Accuracy: 0.65789473684210 Precision:0.5 Recall: 0.846153846153846 F1 score: 0.62857142857142</p> <p>Total Hidden layers: (200,) Accuracy: 0.70940170940170 Precision:0.55263157894736 Recall: 0.552631578947368 F1 score: 0.55263157894736</p>
------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

2)Με δυο κρυφα επιπεδα

<p>Ghost Hidden layers: (50, 25) Accuracy: 0.8 Precision:0.75 Recall: 0.5 F1 score: 0.6</p> <p>Goblin Hidden layers: (50, 25) Accuracy: 0.57894736842105 Precision:0.333333333333333 Recall: 0.230769230769230 F1 score: 0.27272727272727</p> <p>Ghoul Hidden layers: (50, 25) Accuracy: 0.52631578947368 Precision:0.38095238095238 Recall: 0.615384615384615 F1 score: 0.47058823529411</p> <p>Total Hidden layers: (50, 25) Accuracy: 0.63793103448275 Precision:0.44736842105263 Recall: 0.447368421052631 F1 score: 0.44736842105263</p>	<p>Ghost Hidden layers: (100, 50) Accuracy: 0.9 Precision:0.833333333333333 Recall: 0.833333333333333 F1 score: 0.833333333333333</p> <p>Goblin Hidden layers: (100, 50) Accuracy: 0.73684210526315 Precision:0.8 Recall: 0.307692307692307 F1 score: 0.444444444444444</p> <p>Ghoul Hidden layers: (100, 50) Accuracy: 0.78947368421052 Precision:0.619047619047619 Recall: 1.0 F1 score: 0.764705882352941</p> <p>Total Hidden layers: (100, 50) Accuracy: 0.81034482758620 Precision:0.71052631578947 Recall: 0.710526315789473 F1 score: 0.71052631578947</p>	<p>Ghost Hidden layers: (200, 100) Accuracy: 0.875 Precision:0.81818181818181 Recall: 0.75 F1 score: 0.78260869565217</p> <p>Goblin Hidden layers: (200, 100) Accuracy: 0.65789473684210 Precision:0.5 Recall: 0.307692307692307 F1 score: 0.38095238095238</p> <p>Ghoul Hidden layers: (200, 100) Accuracy: 0.68421052631578 Precision:0.52631578947368 Recall: 0.769230769230769 F1 score: 0.625</p> <p>Total Hidden layers: (200, 100) Accuracy: 0.74137931034482 Precision:0.60526315789473 Recall: 0.60526315789473 F1 score: 0.60526315789473</p>
-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Παρατηρούμε ότι τα καλύτερα αποτελέσματα τα έχει το νευρωνικό επιπεδο με πρωτο κρυφο επιπεδο 100 και δευτερο 50 με Accuracy = 0.8103448275862069

SVM Μέθοδος:

Τα δεδομένα χωρίζονται σε όμοιους πληθυσμούς για να μπορέσουν αργότερα να δημιουργηθούν οι κλάσεις και ορίζονται συντεταγμένες, τοποθετώντας γραμμές σαν σύνορα για να χωρίσουν τα δεδομένα με την μικρότερη απόσταση από το πιο κοντινό αντικείμενο του κάθε group. Δημιουργούμε ένα πίνακα με δυο ειδών kernel (linear και rbf). Φορτώνουμε τα δεδομένα στο svm μοντέλο, το εκπαιδεύουμε και ακολουθούμε πάλι την ίδια διαδικασία με confusion matrix όπως στις παραπάνω μεθόδους

Τα αποτελέσματα της SVM μεθόδου:

Linear kernel	RBF
Ghost Kernel: linear Accuracy: 0.9473684210526315 Precision: 1.0 Recall: 0.8333333333333334 F1 score: 0.9090909090909091	Ghost Kernel: rbf Accuracy: 0.9 Precision: 0.7857142857142857 Recall: 0.9166666666666666 F1 score: 0.8461538461538461
Goblin Kernel: linear Accuracy: 0.8157894736842105 Precision: 0.75 Recall: 0.6923076923076923 F1 score: 0.7199999999999999	Goblin Kernel: rbf Accuracy: 0.8157894736842105 Precision: 0.75 Recall: 0.6923076923076923 F1 score: 0.7199999999999999
Ghoul Kernel: linear Accuracy: 0.868421052631579 Precision: 0.75 Recall: 0.9230769230769231 F1 score: 0.8275862068965517	Ghoul Kernel: rbf Accuracy: 0.868421052631579 Precision: 0.8333333333333334 Recall: 0.7692307692307693 F1 score: 0.8
Total Kernel: linear Accuracy: 0.8771929824561403 Precision: 0.8157894736842105 Recall: 0.8157894736842105 F1 score: 0.8157894736842104	Total Kernel: rbf Accuracy: 0.8620689655172413 Precision: 0.7894736842105263 Recall: 0.7894736842105263 F1 score: 0.7894736842105263

Παρατηρούμε ότι το καλύτερο accuracy το πετυχαίνουμε με linear kernel και συγκεκριμένα είναι Accuracy = 0.8771929824561403

Naïve Bayes Μεθοδος:

Είναι ένας αλγόριθμος για classification ο οποίος βασίζεται σύμφωνα με το θεώρημα του Bayes στη μοναδικότητα των δεδομένων. Δηλαδή στο ότι κάποιες ιδιότητες και χαρακτηριστικά είναι μοναδικά σε συγκεκριμένα αντικείμενα και μέσα από αυτά μπορούμε να καθορίσουμε το είδος και να τα διαχωρίσουμε

Περνουμε τα πρωτα τεσσερα features και τα ταξινομουμε με την κατανομη normal distribution(gaussian) και το color το κατανομουμε με multinomial distribution.Επειτα βρισκουμε το γινομενο των δυο πινακων (βρηκαμε στο github μια βιβλιοθηκη για mixed naïve bayes κωδικα που ταιριαζει στην ασκηση μας δηλαδη με πολλαπλασιασμο multi με gaus και το χρησιμοποιησαμε για να βρουμε τα αποτελεσματα) επειτα τα κανονικοποιησαμε κανουμε τις προβλεψεις και confusion matrix όπως στις παραπνω μεθοδους και βρισκουμε τα τελικα αποτελεσματα

Τα αποτελεσματα για Naïve Bayes μεθοδο:

```
Ghost
Accuracy: 0.8717948717948718
Precision:0.7692307692307693
Recall: 0.8333333333333334
F1 score: 0.8

Goblin
Accuracy: 0.8157894736842105
Precision:0.875
Recall: 0.5384615384615384
F1 score: 0.6666666666666667

Ghoul
MultinomialNB
Accuracy: 0.7894736842105263
Precision:0.6470588235294118
Recall: 0.8461538461538461
F1 score: 0.7333333333333334

Total
Accuracy: 0.8260869565217391
Precision:0.7368421052631579
Recall: 0.7368421052631579
F1 score: 0.7368421052631579
```

Παρατηρουμε ότι το Accuracy της Naive Bayes μεθοδου είναι Accuracy= 0,8260869565217391

Εν κατακλείδι το καλύτερο Accuracy το πήραμε από την SVM μέθοδο και συγκεκριμένα με linear kernel (Accuracy = 0.8771929824561403). Η εξήγηση της απόδοσης αυτής που μπορεί να δοθεί είναι η γραμμικότητα του μοντέλου (linear) και στην λειτουργία της SVM μεθόδου που με την χρήση επιτυγχάνονται αποτελέσματα υψηλής απόδοσης