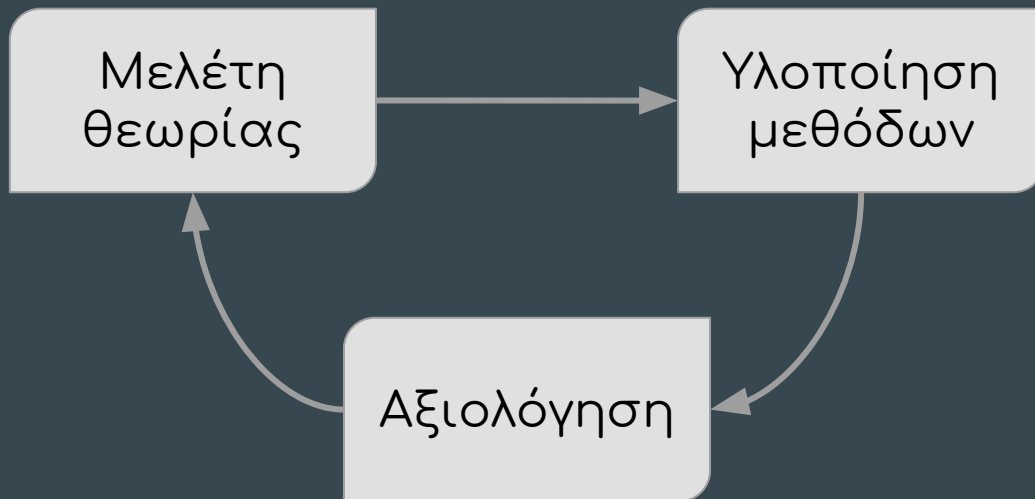


Ανάπτυξη βιβλιοθήκης λογισμικού μηχανικής μάθησης για παράλληλα υπολογιστικά συστήματα



Ανδρέας Μπαμπούρης

Τρόπος εργασίας



Η βιβλιοθήκη «metis»



Ανάπτυξη σε C++



Χρήση βιβλιοθήκης Eigen
για γραμμική άλγεβρα



Παραλληλισμός μέσω
προτύπου OpenMP

Παράλληλη υπολογιστική

OpenMP

→ Συναρτήσεις βιβλιοθήκης

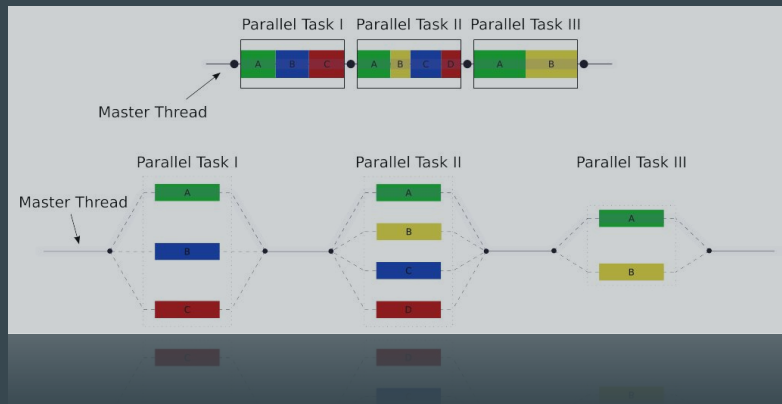
- ◆ `omp_set_num_threads()`
- ◆ `omp_get_thread_num()`

→ Οδηγίες παραλληλοποίησης

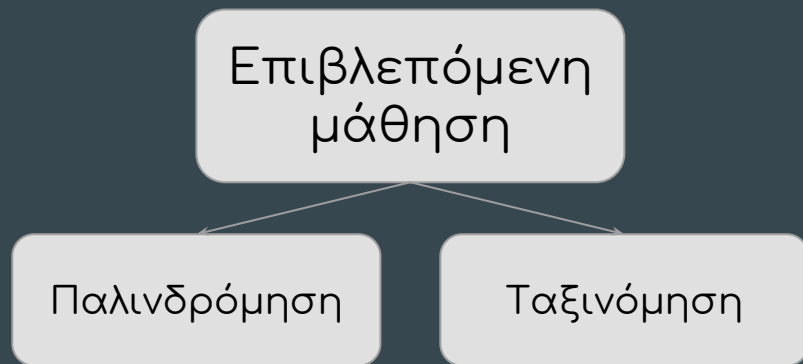
- ◆ `#pragma omp parallel`
- ◆ `#pragma omp for`

→ Οδηγίες συγχρονισμού

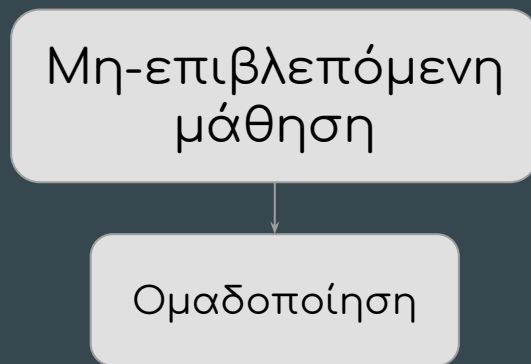
- ◆ `#pragma omp critical`
- ◆ `#pragma omp barrier`



Μηχανική μάθηση



```
void fit(DataLabeled *data);  
VectorXi predict(DataSet *data);  
double score(DataLabeled *data);
```



```
VectorXi cluster(DataSet *data);  
VectorXi predict(DataSet *data);  
double score(DataSet *data);
```

Επεξεργασία δεδομένων → Διαδικασία μάθησης → Επιλογή μοντέλου

Δεδομένα



```
std::string filePath = "../data/iris.data.txt";  
metis::DataSet *input = new metis::DataSet(filePath, ',');  
input->create({0,1,2,3});  
input->applyStandardization();
```

DataSet

- Ανάγνωση
- Αποθήκευσή
- Προεπεξεργασία
- Μαθηματικές πράξεις

Γραμμική παλινδρόμηση

LinearRegression

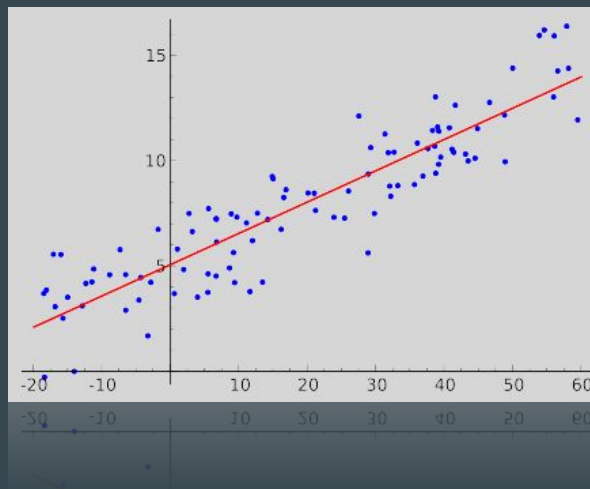
→ Γραμμικό μοντέλο

$$\hat{y} = \theta_0 + \mathbf{X}\hat{\theta}$$

→ Προσδιορισμός των παραμέτρων

→ Μέθοδος ελαχίστων τετραγώνων

```
_coeff.col(o) = (in->transpose() * (*in)).inverse()  
               * (in->transpose()) * out->col(o);  
_intercept.coefRef(o) = out->col(o).mean();
```



Λογιστική παλινδρόμηση

LogisticRegression

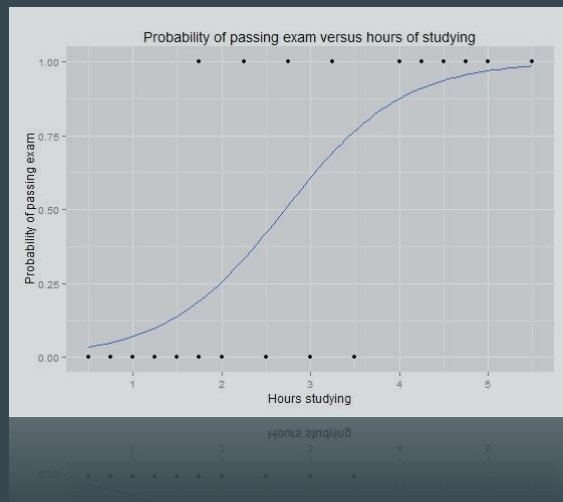
→ Γραμμικός
μετασχηματισμός εισόδου
→

→ Σιγμοειδής συνάρτηση

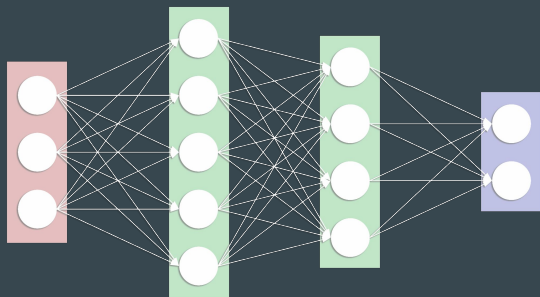
$$S(x) = \frac{1}{1 + e^{-x}}$$

→ Μέθοδος βελτιστοποίησης
επικλινούς καθόδου

```
#pragma omp parallel for
for (unsigned c = 0; c < _nClasses; ++c) {
    if (_batchSize == 1) stochasticGradientDescent(c);
    else batchGradientDescent(c);
}
```



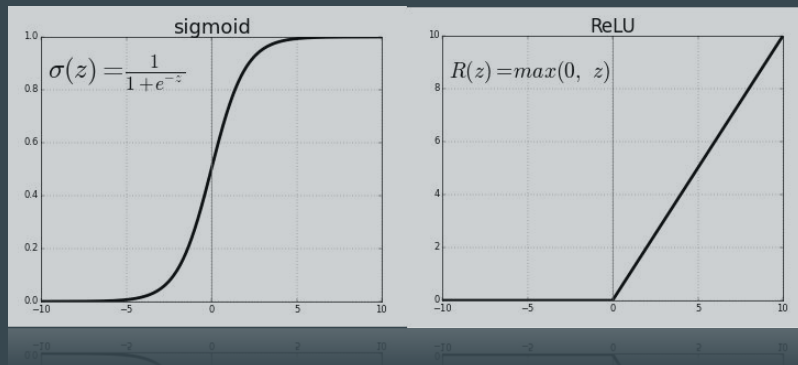
Νευρωνικά δίκτυα



Επίπεδο εισόδου

Κρυφά επίπεδα

Επίπεδο εξόδου



→ Μεταβλητές ΤΝΔ:

- ◆ μέγεθος
- ◆ τρόπος διασύνδεσης
- ◆ συνάρτηση ενεργοποίησης

→ Λειτουργία νευρώνα \approx λογιστικού παλινδρομητή

→ Έξοδος ενός γίνεται είσοδος για έναν άλλο

```
void activationSigmoid(MatrixXd *mat);  
void activationReLU(MatrixXd *mat);
```

Νευρωνικά δίκτυα - Multilayer perceptron

MLPClassifier

- Ρύθμιση της αρχιτεκτονικής
- Επιλογής μεγέθους mini-batch και νημάτων εκτέλεσης
- Υπολογισμός προβλέψεων με εμπρόσθια τροφοδότηση
- Εκπαίδευση μέσω οπισθοδρομικής διάδοσης

```
#pragma omp parallel reduction(+:g_counter)
{
    if (_batchSize == 1) stochasticGradientDescent();
    else batchGradientDescent();
};
```

```
for (unsigned b = th; b < g_nBatches; b += g_nThreads)
{
    // Feed-forward
    linear[0] = g_coeff[th][0] * g_in->block(b *
        _batchSize, 0, _batchSize, _nAttributes).transpose();
    linear[0].colwise() += g_intercept[th][0];
    activated[0] = linear[0];
    activationFunction(&activated[0], _activation[0]);
    // ...
}
```

```
#pragma omp barrier
    if (th == 0) {
        if (_verbose) combineWeights();
    }
#pragma omp barrier
```

Απλός ταξινομητής Bayes

```
for (unsigned a = 0; a < _nAttributes; ++a)
    probabilities.array().row(c) *= findLikelihood(a,
data, c).transpose().array();
probabilities.array().row(c) *= _prior.coeff(c);
```

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

Diagram labels:

- Likelihood: points to $P(x|c)$
- Class Prior Probability: points to $P(c)$
- Posterior Probability: points to $P(c|x)$
- Predictor Prior Probability: points to $P(x)$

$$P(c|X) = P(x_1|c) \times P(x_2|c) \times \dots \times P(x_n|c) \times P(c)$$

$$b(c|X) = b(x^1|c) \times b(x^2|c) \times \dots \times b(x^N|c) \times b(c)$$

Naive Bayes

- Βασίζεται στο νόμο του Bayes
- Υπόθεση ανεξαρτησίας μεταξύ μεταβλητών
- Έλλειψη επαναληπτικού βήματος

Απλός ταξινομητής Bayes - Multinomial & Gaussian

MultinomialNB

- Διαχείριση ονομαστικών τιμών στις εισόδους
- Εκπαίδευση μέσω υπολογισμού συχνοτήτων εμφάνισης τιμών

```
if (zfp)
  for (unsigned a = 0; a < _nAttributes; ++a)
    _likelihood[a].array() += 1.0;
```

GaussianNB

- Διαχείριση αριθμητικών τιμών στις εισόδους
- Εύρεση μέσου όρου και τυπικής απόκλισης για κάθε χαρακτηριστικό

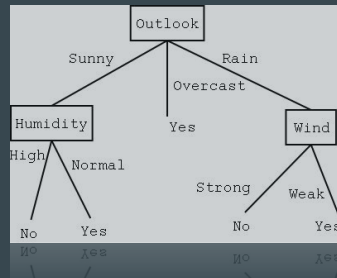
$$P(x = u|c_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} e^{-\frac{(u-\mu)^2}{2\sigma_k^2}}$$

Εκμάθηση δέντρων απόφασης

DecisionTree

- Κάθε κόμβος είναι ερώτημα για ένα χαρακτηριστικό
- Έκβαση στα φύλλα
- Διχοτόμηση κόμβων βάσει τιμών ενός χαρακτηριστικού
- ID3: Μέτρηση κέρδους πληροφορίας για επιλογή

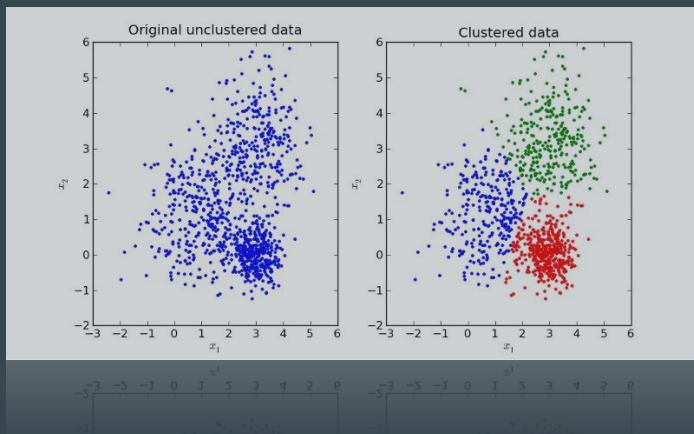
```
for (unsigned c = 0; c < _nChildren; ++c)
    if (data.coeff(_splitAttribute) == _splitValues.coeff(c))
        return _children[c]->traverse(data);
```



```
for (unsigned a = 1; a < attributes.size(); ++a) {
    infoGain = measureInfoGain(data, subSet, attributes[a],
    &newSubSets, &newEntropies);
    if (infoGain >= bestGain) {
        bestGain = infoGain;
        bestSplit = attributes[a];
        *bestSubSets = newSubSets;
        *bestEntropies = newEntropies;
    }
}
```

Ομαδοποίηση k -μέσων

```
#pragma omp for
for (unsigned i = 0; i < _nInstances; ++i) {
    selectedCluster = closestCentroid(g_data->row(i));
    g_newCentroids[th].row(selectedCluster) +=
g_data->row(i);
    g_newClusterSizes[th][selectedCluster] += 1;
}
```



KMeans

- Διαχωρισμός χώρου σε k υποχώρους
- Επιλογή k αρχικών κεντροειδών
- Επαναπροσδιορισμός ομάδων και κέντρων μέχρι σύγκλιση
- Χρήση της μεθόδου αγκώνα για εύρεση του κατάλληλου k

Συνεχής ανάπτυξη της βιβλιοθήκης

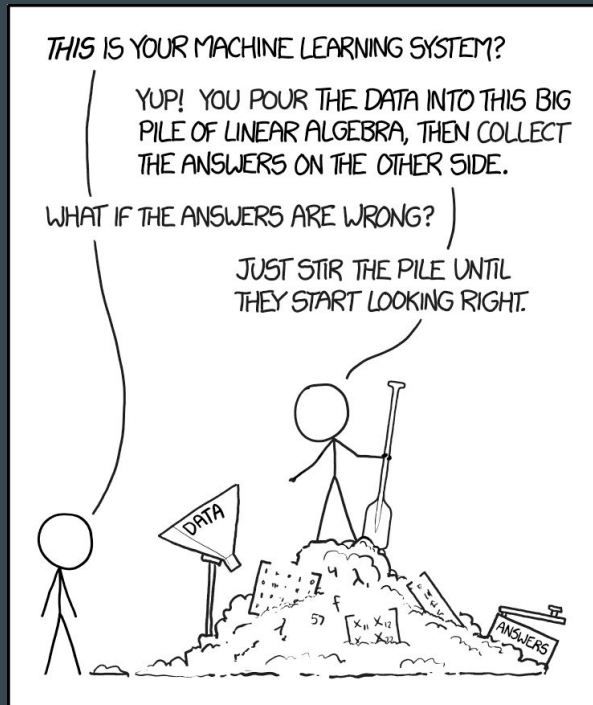
- Παροχή της βιβλιοθήκης «metis» ως ανοιχτό λογισμικό
- Συνέχιση της ανάπτυξης με περισσότερες μεθόδους και πιο αποδοτικές υλοποιήσεις
- <https://github.com/andbamp/metis>

The GitHub logo, featuring the word "GitHub" in a bold, black, sans-serif font. The letters are slightly shadowed, giving it a 3D appearance as if it's floating or attached to a surface.

Συμπεράσματα

- Υλοποίηση μεθόδων: συμβάλλει σημαντικά στην κατανόησή τους
- Ανάδραση μεταξύ αποτελεσμάτων και λεπτομερειών υλοποίησης
- Προεπεξεργασία των δεδομένων: εξαιρετικής σημασίας
- Φυσικός παραλληλισμός σε πολλές μεθόδους μηχανικής μάθησης
- Τεράστια περιθώρια περαιτέρω ανάπτυξης

Ευχαριστώ!



<https://xkcd.com/1838/>