

2η Εργασία στο μάθημα
Νευρωνικά Δίκτυα
Κατηγοριοποίηση του προβλήματος της
MNIST με Support Vector Machine
διαφόρων παραμέτρων.

Δημητριάδης Σπύρος

Αξιολόγηση του SVM στην αναγνώριση χειρόγραφων ψηφίων, τα δεδομένα πάρθηκαν από την MNIST data set. Αρχικά θα παραθέσω κάποιες πληροφορίες για τον αλγόριθμο του SVM που χρησιμοποιήθηκε. Έπειτα αναφέρονται τα αποτελέσματα με διαφορετικά είδη πυρίνων και με διαφορετικές τιμές των παραμέτρων τους. Τέλος υπάρχει συγκρισή του πιο αποδοτικού SVM με την κατηγοριοποίηση 1 και 3 πλησιέστερου γείτονα (Nearest Neighbor).

Ο αλγόριθμος:

Support Vector Machine (SVM)

Για δι-διάστατο (binary) προβλήματα ο SVM κατασκευάζει ένα υπερεπίπεδο για να διαχωρίσει τις δύο κλάσεις, έτσι ώστε η απόσταση του διαχωρισμού (margin) μεταξύ δύο κλάσεων να εγιστοποιείται. Τα διανύσματα των πιο κοντινών στοιχείων στο υπερεπίπεδο αυτό είναι τα υποστηρικτικά διανύσματα (support vectors). Αν έχουμε n παραδείγματα για εκπαίδευση (training samples) και κάθε ένα το συμβολίζουμε με x_i με κλάση y_i , όπου $y_i \in \{-1, 1\}$. Για μη-γραμμική ταξινόμηση εφαρμόζουμε το kernel trick για να μεγιστοποιηθεί το υπερεπίπεδο της απόστασης διαχωρισμού των κλάσεων. Ο αλγόριθμος είναι ίδιος με εκείνον του γραμμικού ταξινομητή, αλλάζουμε όμως το εσωτερικό γινόμενο με έναν μια συνάρτηση μη-γραμμικού πυρίνα. Επομένως αυτή η μεταλλαγή ανεβάζει το πρόβλημα σε μεγαλύτερη διάσταση όπου ο διαχωρισμός στο χώρο εισόδου μπορεί να μην είναι πλέον γραμμικός. Κάποιοι σύνηθες πυρήνες είναι:

- Linear Kernel : $K(x, x') = x \cdot x'$
- Polynomial Kernel : $K(x, x') = (1 + x \cdot x')^p$
- Radial Basis Kernel : $K(x, x') = \exp(-\gamma \|x \cdot x'\|^2)$

Ο πυρίνας συνδέεται με την $\phi(x)$ με την ισότητα $K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$, όπου $\phi(x)$ είναι μια συνάρτηση που αντιστοιχεί την είσοδο x_i σε ένα χαρακτηριστικό διάνυσμα (feature vector). Επομένως ο τύπος του προβλήματος βελτιστοποίησης παίρνει την μορφή:

$$\max \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j [\phi(x_i) \cdot \phi(x_j)]$$

$$\text{s.t. } 0 \leq \alpha_i \leq C, i = 1, \dots, n, \sum_{i=1}^n \alpha_i y_i = 0.$$

Διαφορετικοί πυρίνες έχουν και διαφορετικές επιδόσεις. Για κατηγοριοποίηση πολλών κλάσεων (multi-class classification), το πρόβλημα μετατρέπεται σε κατηγοριοποίηση δυο κλάσεων με τη μέθοδο ένας εναντίων όλων (one-versus-all) ή ένας εναντίων έναν (one-versus-one).

Συμπεράσματα και αποτελέσματα:

Χρησιμοποιήθηκε η `libsvm` για την εκπαίδευση και τον έλεγχο του ταξινομητή SVM. Εξετάστηκε το ίδιο πρόβλημα με 3 διαφορετικούς πυρίνες. Στον Πίνακα 1 βλέπουμε αναλυτικά τα αποτελέσματα.

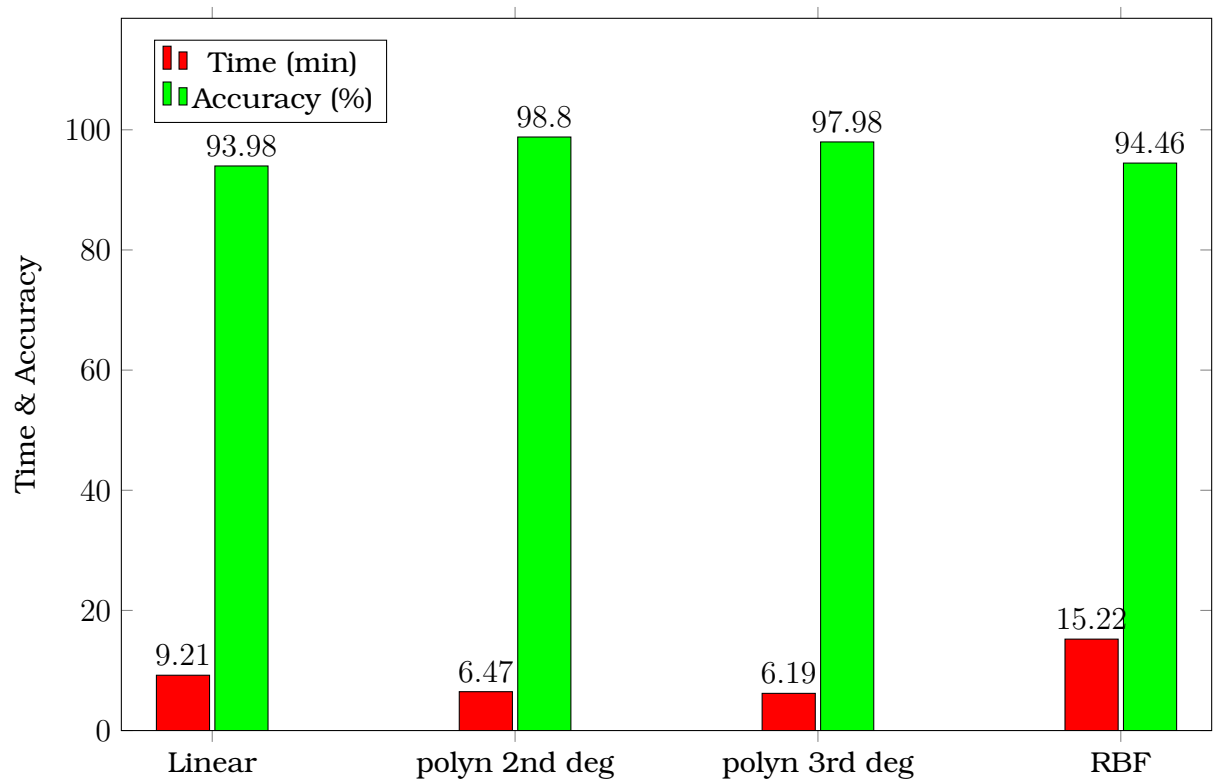
| Kernel Type | Training accuracy (%) | Testing accuracy (%) | Time (min) |
|---------------|-----------------------|----------------------|------------|
| Linear | 97.98 | 93.98 | 9.21 |
| polyn 2nd deg | 100 | 98.08 | 6.472 |
| polyn 3rd deg | 100 | 97.98 | 6.193 |
| RBF | 100 | 94.46 | 15.22 |

Πίνακας 1: Πίνακας με τα ποσοστά επιτυχίας στα στάδια εκπαίδευσης και ελέγχου και ο χρόνος εκμάθησης.

Παρατηρούμε ότι μόνο ο γραμμικός πυρίνας δεν έχει 100% επιτυχία στο στάδιο της εκπαίδευσης. Επίσης ο πολυωνυμικός πυρίνας έχει τις καλύτερες επιδόσεις, με το μεγαλύτερο ποσοστό 98.08% όταν είναι 2ου βαθμού και με χρόνο 6.193 min. Ο πιο χρονοβόρος πυρίνας ήταν ο RBF με χρόνο 15.22 min έχοντας και χαμηλό ποσοστό επιτυχία, 94.46%. Συγκεκριμένα για τον RBF kernel έγιναν πολλές δοκιμές, αλλάζοντας τη τιμή της παραμέτρου γ . Αρχικά πήρα για $\gamma = 1$ και παρατήρησα ότι έκανε πολύ χρόνο χωρίς μεγάλο ποσοστό επιτυχίας. Έτσι το άλλαξα σε $\gamma = \frac{1}{d}$, που είναι η προεπιλογή της `libsvm`, όπου εδώ $d = 784$, και πήρα καλό ποσοστό με πολύ λιγότερο χρόνο.

Χρησιμοποιώντας μη-γραμμικό πυρίνα όπως ο RBF kernel παίρνει περισσότερο χρόνο εκπαίδευσης διότι χρειάζεται να υπολογίσει το γινόμενο πυρίνων για όλα τα υποστηρικτικά διανύσματα, έχοντας όμως περισσότερη επιτυχία από τον γραμμικό. Πράγματι παρατηρούμε ότι ο πολυωνυμικός πυρίνας 3ου βαθμού έχει 7838 support vectors το μικρότερο αριθμό από τους 4. Τα υποστηρικτικά διανύσματα των διαφορετικών πυρίνων φαίνονται στον παρακάτω πίνακα.

| Kernel Type | Number of Support Vectors |
|--------------------|---------------------------|
| Linear | 10347 |
| polynomial 2nd deg | 8729 |
| polynomial 3rd deg | 7838 |
| RBF | 19625 |



Σχήμα 1: Χρόνος και ποσοστό επιτυχίας διάφορων kernels.

Σύναψη:

Τέλος, έχοντας από προηγούμενη εργασία τα αποτελέσματα της κατηγοριοποίησης 1ου και 3ου πλησιέστερου γείτονα συμπεραίνουμε ότι ο χρόνος του πιο αποδοτικού πλησιέστερου γείτονα είναι πολύ περισσότερος σε σχέση με το πιο αργό SVM. Αν συγκρίνουμε την καλύτερη επίδοση του SVM 98.08% με αυτή του Nearest Neighbor 97.05% βλέπουμε ότι ο SVM έχει 1% μεγαλύτερο ποσοστό επιτυχίας και ο χρόνος του είναι αισθητά μικρότερος, περίπου 1.5 ώρα. Επομένως μπορούμε να συμπεράνουμε ότι στο πειράμα μας ο SVM δουλεύει καλύτερα και γρηγορότερα σε σχέση με τη κατηγοριοποίηση πλησιέστερου γείτονα.