ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ Η/Υ ΚΑΙ ΠΛΗΡΟΦΟΡΙΚΗΣ
ΕΡΓΑΣΤΗΡΙΟ ΕΠΕΞΕΡΓΑΣΙΑΣ ΣΗΜΑΤΩΝ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

Εφαρμογες Της Ψηφιακης Επεξεργασιας Σηματων

Εκμαθηση Λεξικου

Αραιη Κωδικοποιηση

Διδάσκων: Αναπλ. Καθηγητής Εμμανουήλ Ζ. Ψαράκης
Επικουρικό έργο: Ευάγγελος Σαρτίνας, Παναγιώτης Γεωργαντόπουλος

Πάτρα Οκτώβριος 2021

Εκμαθηση Λεξικου

Ως λεξικό ορίζουμε μία συλλογή από στοιχειώδη δομικά στοιχεία με τα οποία μπορούμε να συνθέτουμε σύνθετα. Σε παραλληλισμό με την γραμμική άλγεβρα, τα στοιχειώδη αυτά δομικά στοιχεία είναι διανύσματα τα οποία δεν είναι ορθογώνια απαραίτητα μεταξύ τους και το πλήθος τους είναι κατά πολύ μεγαλύτερο από τον ελάχιστο αριθμό που απαιτεί ο ορισμός μίας βάσης του διανυσματικού χώρου (υπερπλήρης βάση), με γραμμικούς συνδυασμούς των οποίων μπορούμε να συνθέσουμε άλλα διανύσματα.

Συνήθως, τα δομικά στοιχεία του γραμμικού συνδυασμού υπακούουν κάποιους περιορισμούς, για να έχει νόημα και η ανάλυση. Στην συγκεκριμένη περίπτωση θα θέσουμε τον περιορισμό να μην χρησιμοποιούνται πάνω από S δομικά στοιχεία για την αναπαράσταση του τελικού διανύσματος.
Έστω $\{\mathbf{d}_k\}_{k=1}^{K}$ με $\mathbf{d}_k \in \mathbb{R}^n$, ένα σύνολο τέτοιων διανυσμάτων και έστω $\{x_k\}_{k=1}^{K}$ με $x_k \in \mathbb{R}$ έτσι ώστε:

$$\mathbf{y} = \sum_{k=1}^{K} \mathbf{d}_k x_k = \mathbf{d}_k^T \mathbf{x} \tag{1}$$

όπου $\mathbf{y} \in \mathbb{R}^n$. Αυτό μας δίνει την δυνατότητα χρησιμοποιώντας το ίδιο σύνολο $\{\mathbf{d}_k\}_{k=1}^{K}$ να συνθέσουμε εν δυνάμει άπειρα στο πλήθος διανύσματα $\mathbf{y}$, αλλάζοντας τα βάρη του γραμμικού συνδυασμού, δηλαδή τα x_k.
Αντίστροφα, εάν διαθέτουμε το διάνυσμα $\mathbf{y}$, θα μπορούσαμε να το αναλύσουμε όπως υποδεικνύει η Σχέση (1).

Στη γενική περίπτωση, εάν διαθέτουμε ένα μητρώο $Y \in \mathbb{R}^{n\times N}$ από διανύσματα τύπου $\mathbf{y}$ τότε, χρησιμοποιώντας τη Σχέση (1), αυτό μπορεί να κωδικοποιηθεί ως ακολούθως:

$$Y = DX \tag{2}$$

όπου, $X \in \mathbb{R}^{K\times N}$ και η l_0 ψευδονόρμα[1] κάθε στήλης του[2] μητρώου X φράσσεται από πάνω από μία σταθερά $T << K$, δηλαδή $||\mathbf{x}_n||_0 \leq T,\ n = 1, \cdots, N$.

Σε κάθε περίπτωση βλέπουμε πως η Σχέση (2) δεν μπορεί να τηρείται αυστηρά. Για αυτό καταφεύγουμε σε ελαχιστοποίηση της διαφοράς των μελών της, οριζόντας το ακόλουθο πρόβλημα βελτιστοποίησης:

[1]Στάθμη (*norm*) είναι μία συνάρτηση $f(.)$ που απεικονίζει τα στοιχεία ενός διανυσματικού χώρου στους μη αρνητικούς πραγματικούς θετικούς, δηλαδή:

$$f:\ \mathbb{R}^N \rightarrow \mathbb{R}$$

αν $\forall\ \mathbf{a},\ \mathbf{b} \in \mathbb{R}^N$ και $\alpha \in \mathbb{R}$ ικανοποιεί τα ακόλουθα:

1. $f(\mathbf{a}+\mathbf{b}) \leq f(\mathbf{a}) + f(\mathbf{b})$
2. $f(\alpha\mathbf{a}) = |\alpha| f(\mathbf{a})$
3. $f(\mathbf{a}) = 0$ αν και μόνο αν $\mathbf{a} = \mathbf{0}$.

[2]Η l_0 ψευδονόρμα ενός διανύσματος εκφράζει το πλήθος των μη μηδενικών στοιχείων του.

$$\mathcal{P}: \min_{X,\ D}\{Y - DX_F^2\}\ s.t.\ x_{n0} \leq T,\ \ n = 1, \dots, N. \tag{3}$$

όπου $\cdot_F$ η νόρμα $Frobenius$.

Ωστόσο, είναι φανερό ότι απαιτείται η επίλυση δύο ξεχωριστών προβλημάτων τα οποία δεν μπορούμε να λύσουμε ταυτόχρονα. Συγκεκριμένα:

- $\mathcal{P}_1$: Το πρώτο πρόβλημα είναι αυτό της **αραιής κωδικοποίησης** το οποίο μπορεί να διατυπωθεί ως εξής: Δοθέντος ότι τα μητρώα D και Y είναι γνωστά, χρειαζόμαστε το X εκείνο που κάθε στήλη του υπακούει, ως προς την l_0 ψευδονόρμα της, τον περιορισμό της Σχέσης (2). Με άλλα λόγια ψάχνουμε μία αραιή κωδικοποίηση του Y. Ορίζοντας το πρόβλημα πιο αυστηρά, έχουμε:

$$\mathcal{P}_1:\ X^* = \arg\min_X\{Y - DX_F^2\}\ s.t.\ x_{n0} \leq T,\ \ n = 1, \dots, N. \tag{4}$$

- $\mathcal{P}_2$: Το δεύτερο πρόβλημα είναι αυτό της **εκμάθησης λεξικού** το οποίο μπορεί να διατυπωθεί ως εξής: Δοθέντος του μητρώου X να βρεθεί το μητρώο D ώστε να ισχύει η Σχέση (2). Η αυστηρή μαθηματική διατύπωση του προβλήματος είναι η ακόλουθη:

$$\mathcal{P}_2:\ D^* = \arg\min_D\{Y - DX_F^2\}. \tag{5}$$

Για την άσκηση αυτή θα χρησιμοποιήσουμε συγκεκριμένους αλγορίθμους της βιβλιογραφίας για την επίλυση κάθε προβλήματος. Συγκεκριμένα, για:

- την εκμάθηση του λεξικού θα χρησιμοποιηθεί ο αλγόριθμος $KSVD$ [1] και
- για την αραιή κωδικοποίηση θα χρησιμοποιηθεί ο αλγόριθμος $GenOMP$ [2].

Παρακάτω ακουλουθούν οι κώδικες και για τους δύο αλγόριθμους καθώς και ο αλγόριθμος εκμάθησης λεξικού που λύνει το συνδυαστικό πρόβλημα.

Όπως είπαμε το αρχικό πρόβλημα είναι να βρούμε D, X που να ελαχιστοποιούν το σφάλμα $Y - DX_F^2$. Αφού το αναλύσουμε σε δύο ξεχωριστά υποπροβλήματα παίρνουμε τον αλγόριθμο που φαίνεται παρακάτω. Όπως μπορούμε να δούμε παρακάτω ο αλγόριθμος $GenOMP$ ξεκινώντας με ένα λεξικό και ένα διάνυσμα που θέλουν να περιγράψουμε με αυτό το λεξικό, ψάχνει με άπληστο τρόπο ένα υποσύνολο το πολύ T διανυσμάτων του D ώστε να ελαχιστοποιεί το σφάλμα της Σχέσης (4). Η συνθήκη που τερματίζει των αλγόριθμο είναι είτε ότι φτάσαμε το πλήθος διανυσμάτων που θέλουμε (ώστε να θεωρείται η αναπαράσταση αραιή) ή ότι περιγράψαμε αρκετά καλά το διάνυσμα y (δηλαδή το σφάλμα είναι μικρότερο από ένα προκαθοριμένο αριθμό ϵ). Τέλος, ο αλγόριθμος επιστρέφει ένα αραιό διάνυσμα x που κωδικοποιεί το διάνυσμα εισόδου y.

Algorithm 1 Dictionary Learning (DL)

$Input : D \in \mathbb{R}^{n \times K}, Y \in \mathbb{R}^{n \times N}, \epsilon, T_0 \in \mathbb{R}, noEpochs \in \mathbb{Z}^+$
$Output : D^* \in \mathbb{R}^{n \times K},$
$t \leftarrow 1$
for $(epoch = 1 : noEpochs)$ **do**
  for $i = 1 : N$ **do**
    $X(:, i) \leftarrow GenOMP(D, y, T_0, \epsilon)$
  $(D, X) = KSVD(D, X, Y)$
return D, X

Ο αλγόριθμος αυτός έχει δύο μειονεκτήματα: Πρώτον, παγιδεύεται σε τοπικά ελάχιστα. Δηλαδή δεν βρίσκει ντετερμινιστικά την βέλτιστη αραιή αναπαράσταση του διανύσματος y. Δεύτερον, ο χρόνος εκτέλεσης αυξάνεται γραμμικά με το κατώφλιο T. Με άλλα λόγια εάν θέλαμε να αναπαραστήσουμε το διάνυσμα εισόδου y με διπλάσιο αριθμό ατόμων του λεξικού, θα έπρεπε να εκτελέσουμε διπλάσιο αριθμό επαναλήψεων.

Παρατηρήστε, επίσης, πως πριν την επιλογή κάθε ατόμου από το λεξικό, το σφάλμα $\mathbf{r}_t$ ορθογωνοποιείται ως προς το πιο πρόσφατα επιλεγμένο άτομο, ώστε αυτό να μην ξαναεπιλεγεί στην επόμενη επανάληψη. Αυτό επιτυγχάνεται με τη χρήση του μητρώου:

$$P = I - D_{S^t}(D_{S^t}^T D_{S^t})^{-1} D_{S^t}^T$$

το οποίο, όπως εύκολα μπορεί κανείς να αποδείξει, είναι ένα μητρώο ορθογώνιας προβολής.

Algorithm 2 Generalized Orthogonal Matching Pursuit (GenOMP)

$Input : D \in \mathbb{R}^{n \times K}, \mathbf{y} \in \mathbb{R}^n, \epsilon, T_0 \in \mathbb{R}$
$Output : \mathbf{x} \in \mathbb{R}^K$
$\mathbf{r} \leftarrow \mathbf{y}$
$S^0 \leftarrow \emptyset$
while $(\mathbf{r}_2 > \epsilon) \wedge (|S| \leq T_0)$ **do**
  $k^* = \arg\max_k |\mathbf{d}_k^T \mathbf{r}|\}$
  $S \leftarrow S \cup k^*$
  $\mathbf{x}(k^*) \leftarrow \mathbf{d}_{k^*}^T \mathbf{r}$
  $\mathbf{r} \leftarrow (I - D_S(D_S^T D_S)^{-1} D_S^T)\mathbf{r}$
return $\mathbf{x}$

Όσον αφορά την εκμάθηση του λεξικού ο $KSVD$ αναλύει το σφάλμα σε αυτό που οφείλεται σε ένα συγκεκριμένο άτομο του λεξικού και στα υπόλοιπα. Αυτό επιτρέπει να ενημερώνει το λεξικό άτομο προς άτομο με τον υπολογισμό μόνο ενός τάξης 1 μητρώου.

$$\begin{aligned} Y - DX_F^2 &= Y - \sum_{j=1}^{K} \mathbf{d}_j \mathbf{x}_j^t {}_F^2 \\ &= Y - \sum_{j \neq k}^{K} \mathbf{d}_j \mathbf{x}_j^t - \mathbf{d}_k \mathbf{x}_k^t {}_F^2 \\ &= E_k - \mathbf{d}_k \mathbf{x}_k^t{}_F^2 \end{aligned} \tag{6}$$

όπου $\mathbf{x}_k^t$ η k-οστή γραμμή του X.

Σύμφωνα με την Ανάλυση Ιδιάζουσων Τιμών, ένα μητρώο τάξης n μπορεί να αναλυθεί σε άθροισμα n τάξης 1 μητρώων. Δηλαδή:

$$A = V \Sigma U^T = \sum_{i=1}^{n} \sigma_i \mathbf{v}_i \mathbf{u}_i^t$$

Οπότε εάν θέλουμε η Σχέση (6) να ελαχιστοποιείται, εφόσον το E_k είναι σταθερό, μπορούμε να αντικαταστήσουμε όπου $\mathbf{d}_k, \mathbf{x}_k^t$ τα $\mathbf{v}_k$ και $\sigma_k \mathbf{u}_k^t$, αντίστοιχα.

Algorithm 3 KSVD

```
Input : D ∈ ℝ^{n×K}, Y ∈ ℝ^{n×N}, X ∈ ℝ^{K×N}
Output : D ∈ ℝ^{n×K}, X
for i = 1 : K do
    E = Y − DX + D(:, k)X(k, :)
    ind ← X(k, :) not 0
    E_reduced = E(:, ind)
    U, S, V ← SVD(E_reduced)
    (u, s, v) ← (U(:, 1), S(1, 1), V(:, 1))
    D(:, k) ← u
    X(k, :) ← 0
    X(k, ind) ← s v^t
return D, X
```

Τέλος, να αναφέρουμε πως ένα λεξικό έχει σκοπό να μπορεί να περιγράψει δεδομένα σε έναν χώρο χωρίς να χρειάζεται να τον καλύπτει ολόκληρο. Έτσι μπορεί να διατηρεί μόνο την πληροφορία που αφορά τα δεδομένα εκπαίδευσης και να επιτρέπει κωδικοποιήσεις με πολύ λίγα άτομα (χαμηλής τάξης). Αυτό προϋποθέτει ότι ο πληθικός αριθμός των ατόμων είναι μεγαλύτερος από την διάσταση του χώρου στον οποίο βρίσκονται. Δηλαδή, εάν τα άτομα ανήκουν στον $\mathbb{R}^{64}$ τότε διαθέτουμε τουλάχιστον 65 από αυτά. Σε κάθε περίπτωση, επειδή το λεξικό διαμορφώνεται ανάλογα με το είδος των δεδομένων που θέλουμε να αναπαραστασήσουμε, χρειαζόμαστε όσο περισσότερα δείγματα από τα δεδομένα αυτά. Επειδή βασίζουμε την εκπαίδευση του λεξικού στην Ευκλείδια νόρμα, ποτέ δύο εικόνες δεν πρόκειται να είναι αρκετά κοντά ώστε να τις ομαδοποιήσουμε, ακόμη και αν αυτές απεικονίζουν παρόμοιας κλάσης αντικείμενα. Για αυτό καταφεύγουμε σε περιγραφή κομματιών της εικόνας, κάτι το οποίο μας βολεύει τόσο για την διάσταση των δεδομένων όσο και για το πλήθος τους (βλ. Υπόδειξη 2 για περισσότερες πληροφορίες). Το πόσο μικρό ή μεγάλο πρέπει να είναι ένα κομμάτι είναι μέρος μίας συζήτησης που ξεφεύγει από τα όρια αυτής της εργασίας.

ΔΙΑΔΙΚΑΣΙΑ

Στο πλαίσιο αυτής της άσκησης, καλείστε να:

1. Αναπτύξετε στο περιβάλλον $Matlab$ (και $Python(bonus)$), τον αλγόριθμο DL ακριβώς όπως αυτός ορίστηκε παραπάνω. Φροντίστε, για την δική σας ευκολία, να αναπτύξετε ξεχωριστές συναρτήσεις για την υλοποίηση του $GenOMP$ και της $KSVD$.

2. Παρουσιάστε σε γράφημα: το συνολικό μέσο τετραγωνικό σφάλμα $Y - DX_F^2$ για κάθε εποχη, εκπαιδεύοντας το λεξικό στις $training$ εικόνες.

3. ($Bonus$) Παρουσιάστε την αρχική και τελική μορφή του λεξικού D σε μορφή εικόνας. Δηλαδή, κάθε διάνυσμά του θα μετασχηματίζεται σε ένα τετραγωνικό μητρώο. Κάθε τέτοιο μητρώο θα εισάγεται σε ένα μεγαλύτερο μητρώο ώστε να προκύπτει μια οπτικοποιημένη μορφή των ατόμων του λεξικού.

5. Παρουσιάστε το σφάλμα ανακατασκευής για όλες τις $testing$ εικόνες, αφού έχετε τελειώσει με την εκπαίδευση του λεξικού. Σχολιάστε το σφάλμα ως προς το είδος της κάθε $testing$ εικόνας. Εάν θέλετε να χρησιμοποιείσεται δικές σας εικόνες κάντε το. Φροντίστε, ωστόσο, να διατηρείται το μοτίβο των εικόνων που σας δίνονται. Οι $training$ εικόνες θα πρέπει να είναι ομοειδής. Ενώ οι $testing$ θα πρέπει αν περιέχουν μία εικόνα από το τεστινγ και μία εικόνα ίδιας κλάσης με το $training$.

Υποδείξεις:

1. Το λεξικό D και το μητρώο X θα πρέπει να αρχικοποιούνται με τυχαίο τρόπο. Χρησιμοποιείστε την συνάρτηση $randn$ της $Matlab$ και την $numpy.random.rand$ της $Python$.

2. Τα δεδομένα Y (θα τα βρείτε στο αρχείο $dataset.zip$) θα πρέπει να βρίσκονται σε διανυσματική μορφή. Δηλαδή κάθε στήλη του μητρώου Y θα πρέπει να περιέχει μία εικόνα. Χρησιμοποιείστε τη συνάρτηση $reshape$ της $Matlab$ και την $numpy.reshape$ της $Python$ για το σκοπό αυτό. Επειδή, συνήθως, μία εικόνα επιβάλλει πολύ μεγάλη διάσταση στο Y (π.χ. μία εικόνα διάστασης 100×100 αν διανυσματοποιηθεί θα βρίσκεται στο $\mathbb{R}^{10000}$) είναι καλύτερο να χωρίσετε τις εικόνες σε $block$ διάστασης της επιλογής σας (όπως 8×8) και να διανυσματοποιήσετε εκείνα. Έπειτα, μπορείτε να συνθέσετε ένα μητρώο Y με 64 γραμμές και

$\#blocks$ στήλες. Κατ' αυτόν τον τρόπο θα διαθέτετε και σχετικά μικρή διάσταση δεδομένων και πολύ μεγάλο πλήθος από αυτά. (Οι εικόνες που σας δίνονται έχουν διαστάσεις ακαίρεα πολλαπλάσια του 8, για δική σας ευκολία.)

3. Το μητρώο Y θα πρέπει επίσης να κανονικοποιηθεί ώστε κάθε διάνυσμα του να έχει μέτρο 1.
4. Το μητρώο D θα πρέπει επίσης να είναι κανονονικοποιημένο όπως το Y.
5. Προσεξτε οι εικόνες που θα χρησιμοποιηθούν για την εκπαίδευση του λεξικού να μην χρησιμοποιηθούν και για το $testing$ (πλην μίας, όπως αναφέρεται στο Ερώτημα 5).

Βιβλιογραφία

[1] Michal Aharon, Michael Elad, and Alfred Bruckstein. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on signal processing*, 54(11):4311–4322, 2006.

[2] Jian Wang, Seokbeop Kwon, and Byonghyo Shim. Generalized orthogonal matching pursuit. *IEEE Transactions on signal processing*, 60(12):6202–6216, 2012.