Θέμα 1 Ερωτήσεις θεωρίας (20%)

1. Ποια είναι τα πλεονεκτήματα και ποια τα μειονεκτήματα ενός RNN (recurrent neural network) σε σχέση με ενα feedforward neural network;
2. Σε ποιές περιπτώσεις χρησιμοποιείται ο αλγόριθμος EM (Expecation Maximization) και σε ποιές ο αλγόριθμος Maximum Likelihood (ML);
3. Ποιές είναι οι διαφορές μεταξύ των HMM (hidden markov models) και των MM (markov models);
4. Περιγράψτε το πρόβλημα vanishing/exploding gradient. Σε ποιές περιπτώσεις/αρχιτεκτονικές εμφανίζεται και γιατί ; Μπορείτε να προτείνετε κάποιες λύσεις ;

Θέμα 2 HMM (40%)

Σε αυτό το πρόβλημα θα χρησιμοποιήσετε ένα HMM για την αποκωδικοποίηση μιας απλής ακολουθίας DNA. Είναι γνωστό ότι το DNA αναπαρίσταται ως μια σειρά από βάσεις $\{A, C, G, T\}$. Ας υποθέσουμε ότι υπάρχει μια κρυφή μεταβλητή S η οποία ελέγχει τη δημιουργία μιας ακολουθίας DNA. Η S μπορεί να πάρει 2 πιθανές καταστάσεις, $\{S_1, S_2\}$. Θεωρείστε επίσης ένα HMM M με πιθανότητες μετάβασης:

$$P(S_1|S_1) = 0.8, P(S_2|S_1) = 0.2, P(S_1|S_2) = 0.2, P(S_2|S_2) = 0.8,$$

πιθανότητες παρατηρήσεων:

$$P(A|S_1) = 0.4, P(C|S_1) = 0.1, P(G|S_1) = 0.4, P(T|S_1) = 0.1$$

$$P(A|S_2) = 0.1, P(C|S_2) = 0.4, P(G|S_2) = 0.1, P(T|S_2) = 0.4$$

και αρχικές πιθανότητες:

$$P(S_1) = 0.5, P(S_2) = 0.5$$

Έστω ότι παρατηρείτε την ακολουθία $x = CGTC$. Υπολογίστε:

(α) Την $P(x|M)$ με χρήση του αλγορίθμου forward.

(β) Τις ύστερες (a posteriori) πιθανότητες $P(\pi_i = S_1|x, M)$ για $i = 1, \ldots, 4$.

(γ) Το πιθανότερο μονοπάτι κρυφών καταστάσεων με χρήση του αλγορίθμου $Viterbi$.

Θέμα 3 Ταξινομητές (40%)

Δίνονται τα ακόλουθα δείγματα εκπαίδευσης $D_1 = \{0,\ 1.8,\ 2,\ -2.6,\ 7,\ 8,\ 9,\ 11\}$ και $D_2 = \{2.9,\ 4.7,\ 5.5,\ 6.9\}$ από τις κατηγορίες ω_1 και ω_2 αντίστοιχα.

(α) Χρησιμοποιήστε τον αλγόριθμο μεγιστοποίησης της πιθανότητας (maximum likelihood) για να υπολογίσετε τις κατανομές $p(x|\omega_1)$ και $p(x|\omega_2)$ και τις a priori πιθανότητες $p(\omega_1)$, $p(\omega_2)$. Υποθέστε ότι η κατανομή ω_1 είναι της μορφής $p(x|\omega_1) = w_{11}\mathcal{N}(\mu_{11}, 1) + w_{12}\mathcal{N}(\mu_{12}, 1)$ και η κατανομή ω_2 είναι της μορφής $\mathcal{N}(\mu_2, 1)$ όπου $\mathcal{N}(\mu, \sigma^2)$ είναι η Γκαουσιανή κατανομή:

$$\mathcal{N}(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} exp(-\frac{(x-\mu)^2}{2\sigma^2})$$

(β) Υπολογίστε τα διαστήματα απόφασης για τις κατηγορίες ω_1 και ω_2 σύμφωνα με τον κανόνα απόφασης του Bayes και τις κατανομές που υπολογίσατε στο (α).

(γ) Υπολογίστε τα διαστήματα απόφασης για τις κατηγορίες ω_1 και ω_2 σύμφωνα με τον κανόνα απόφασης των 3 κοντινότερων γειτόνων NNR-3.

(δ) Συγκλίνει (σε λύση) ο αλγόριθμος συναρτήσεων διαχωρισμού perceptron για τις κατηγορίες ω_1 και ω_2;

(ε) Δεδομένου ότι οι πραγματικές κατανομές των δύο κατηγοριών είναι

$$p(x|\omega_1) = 0.5\mathcal{N}(3, 1) + 0.5\mathcal{N}(9, 1) \text{ και } p(\omega_1) = \tfrac{2}{3}$$

$$p(x|\omega_2) = \mathcal{N}(5, 1) \text{ και } p(\omega_2) = \tfrac{1}{3}$$

ποιός από τους αλγόριθμους Maximum Likelihood (α), NNR-3 (γ) ελαχιστοποιεί το λάθος ταξινόμησης για το συγκεκριμένο παράδειγμα; Συγκρίνετε το λάθος ταξινόμησης Bayes με το λάθος ταξινόμησης για κάθε έναν από τους αλγορίθμους (συγκρίνοντας τα σημεία απόφασης).

(στ) Έστω ένα σύνολο δειγμάτων D_t, το οποίο προκύπτει από την ένωση των συνόλων δειγμάτων D_1 και D_2. Συγκεκριμένα:

$$D_t = \{0,\ 1.8,\ 2,\ -2.6,\ 7,\ 8,\ 9,\ 11,\ 2.9,\ 4.7,\ 5.5,\ 6.9\}$$

Θεωρείστε ένα πρόβλημα ταξινόμησης δύο κατηγοριών. Εφαρμόστε τον αλγόριθμο k-means χρησιμοποιώντας *ως αρχικά κέντρα τα σημεία* 1.8 και 5.5 αντίστοιχα.