

Στοχαστικές Διεργασίες & Βελτιστοποίηση

Απαντήσεις Θεωρίας Παλαιών Θεμάτων

ΑΚΑΔ. ΕΤΟΣ 2018-2019

Θέμα 1

(α) Η μη επιβλεπόμενη μάθηση αξιοποιείται σε αλγορίθμους μηχανικής μάθησης που αναλύουν ή συσταδοποιούν δεδομένα τα οποία προέρχονται από δείγματα που δε φέρουν κάποια επισημείωση (label). Στόχος είναι μέσω της μίμησης (για παράδειγμα μέσω μέτρων ομοιότητας) ο αλγόριθμος να κατασκευάσει μια γενικευμένη αναπαράσταση των δεδομένων. Αντιθέτως, η επιβλεπόμενη μάθηση αξιοποιεί δεδομένα με επισημειώσεις με σκοπό την εκπαίδευση αλγορίθμων που να προβλέπουν με όσο το δυνατό περισσότερη ακρίβεια μπορούν μελλοντικά δεδομένα τα οποία δεν είναι επισημειωμένα. Μια άλλη κατηγορία μάθησης είναι η ενισχυτική μάθηση, όπου η διαδικασία μάθησης αφορά πράκτορες (agents) οι οποίοι αλληλεπιδρούν με το περιβάλλον τους με στόχο να μεγιστοποιήσουν τις ανταμοιβές (rewards) που λαμβάνουν για τις διάφορες κινήσεις τους. Σε αντίθεση με την επιβλεπόμενη μάθηση, οι πράκτορες δεν καθοδηγούνται από σαφώς ορισμένες πληροφορίες (π.χ. επισημειωμένα δεδομένα), αλλά καλούνται να βρουν μια ισορροπία ανάμεσα στην εξερεύνηση του περιβάλλοντος στο οποίο δρουν, με στόχο την ανακάλυψη νέας πληροφορίας, και την αξιοποίηση της ήδη υπάρχουσας πληροφορίας προς τη μεγιστοποίηση των rewards τους.

- Supervised Learning: Multi Layer Perceptron, Radial Basis Function Neural Network, Random Forests, LSTMs, Hopfield Networks, Autoencoders

- Unsupervised Learning: k-Means, PCA, Hierarchical Clustering, Gaussian Mixture Models

- Reinforcement Learning: Q-learning, Value Iteration, Policy Iteration, Bellman-Ford

(β) Η κυριότερη χρησιμότητα της μεθόδου PCA είναι το γεγονός πως πραγματοποιεί διαστατική μείωση στο δείγμα, αφού επιτυγχάνει την περιγραφή ενός προβλήματος N αρχικών συνιστωσών προβάλλοντάς τες σε ένα χώρο $M < N$ συνιστωσών, διατηρώντας όμως με ελεγχόμενο τρόπο ένα υψηλό ποσοστό της αρχικής διακύμανσης των δεδομένων (δηλαδή της πληροφορίας που διατηρείται κατά την προβολή).

(γ) Ο περιορισμός (restriction) στον οποίο οφείλεται και το όνομα των RBMs είναι πως το δίκτυο (γράφος) που τις περιγράφει δεν επιτρέπει συναπτικά βάρη ανάμεσα σε νευρώνες που ανήκουν στο ίδιο επίπεδο. Με άλλα λόγια, δεν επιτρέπονται μεταβάσεις από hidden σε hidden και από visible σε visible νευρώνα.

(δ) Σε επίπεδο πιθανοτήτων, δεδομένου ενός δείγματος x , το οποίο ταξινομείται σε ένα σύνολο κατηγοριών y , ένα παραγωγικό (generative) μοντέλο προσπαθεί να υπολογίσει την από κοινού πυκνότητα πιθανότητας $p(x,y)$, ενώ ένα διακριτικό (discriminative) μοντέλο προσπαθεί να υπολογίσει την υπό συνθήκη κατανομή πιθανότητας $p(y|x)$. Διαισθητικά, θα μπορούσε κανείς να πει πως ένα διακριτικό μοντέλο περιορίζεται στο να ταξινομήσει ορθά ένα δειγματικό σημείο εντοπίζοντας τις βασικές διαφορές ανάμεσα στις διάφορες κατηγορίες. Από την άλλη, ένα παραγωγικό μοντέλο προσπαθεί να αντλήσει πληροφορίες για την ίδια την κατανομή που γενικά κάθε κατηγορία, προκειμένου να ταξινομήσει έτσι το δειγματικό σημείο.

Θέμα 3

(α) Ο αλγόριθμος Hill Climbing είναι δεσμευμένος να πραγματοποιεί βήματα τα οποία οδηγούν το σύστημα σε μικρότερη ενέργεια και επομένως είναι επιρρεπής στην παγίδευση σε τοπικά ελάχιστα. Όσο υψηλότερη είναι η θερμοκρασία, τόσο περισσότερο ο αλγόριθμος Simulated Annealing επιτρέπει την εξερεύνηση καταστάσεων ακόμα και μεγαλύτερης ενέργειας, προκειμένου να ελαχιστοποιήσει μακροπρόθεσμα την ενέργεια, φτάνοντας σε κάποιο σημείο το οποίο ενδεχομένως να είναι ολικό και όχι τοπικό ελάχιστο. Στο όριο $T \rightarrow 0$, οι δύο αλγόριθμοι ταυτίζονται.

ΑΚΑΔ. ΕΤΟΣ 2019-2020

Θέμα 1

Supervised Learning: Random Forests, Multilayer Perceptron, Radial Basis Function (RBF) Neural Network

Unsupervised Learning: k-means Clustering, Principal Component Analysis

Reinforcement Learning: Q-Learning, Value Iteration, Bellman-Ford

Σημείωση: Το RBF NN περιλαμβάνει συχνά ένα κομμάτι *unsupervised learning*, καθώς είθισται στην αρχικοποίησή του να χρησιμοποιείται ο *k-means* για τον προσδιορισμό των κέντρων.

Θέμα 2

(α) Η κύρια διαφορά ανάμεσα στην ενισχυτική μάθηση και την επιβλεπόμενη μάθηση είναι η διαδικασία με την οποία πραγματοποιείται η εκπαίδευση του εκάστοτε αλγορίθμου. Συγκεκριμένα, κατά την επιβλεπόμενη μάθηση, το μοντέλο καλείται να μάθει να προβλέπει τιμές ή να ταξινομεί, βάσει επισημειώσεων (labels) οι οποίες συνοδεύουν τα διανύσματα χαρακτηριστικών των δειγματικών σημείων. Από την άλλη, στην ενισχυτική μάθηση, εκπαιδεύεται ένας πράκτορας (agent) μέσω αλληλεπίδρασής του με το περιβάλλον του. Κάθε του πράξη ή ακολουθία αυτών έχει ως αποτέλεσμα την αποζημίωση ή την τιμωρία του, επομένως η εκμάθησή του δε βασίζεται σε ένα προκαθορισμένο κανόνα μέσω επισημειώσεων, αλλά μέσω της διερεύνησης διαφόρων πρακτικών, προκειμένου να βρει κάποια που να του αποφέρει το μέγιστο κέρδος. (βλ. *επίσης* Θέμα 1, 2019).

(β) Παραθέτουμε αρχικά μια εξαιρετική απάντηση από το StackOverflow:

In policy iteration algorithms, you start with a random policy, then find the value function of that policy (policy evaluation step), then find a new (improved) policy based on the previous value function, and so on. In this process, each policy is guaranteed to be a strict improvement over the previous one (unless it is already optimal). Given a policy, its value function can be obtained using the Bellman operator.

In value iteration, you start with a random value function and then find a new (improved) value function in an iterative process, until reaching the optimal value function. Notice that you can derive easily the optimal policy from the optimal value function. This process is based on the optimality Bellman operator.

In some sense, both algorithms share the same working principle, and they can be seen as two cases of the generalized policy iteration. However, the optimality Bellman operator contains a max operator, which is non linear and, therefore, it has different features. In addition, it's possible to use hybrid methods between pure value iteration and pure policy iteration.

Πώς θα απαντούσαμε:

Στα πλαίσια του αλγορίθμου Policy Iteration, ξεκινά κανείς με μια τυχαία πολιτική και υπολογίζει την αντίστοιχη value function της. Στη συνέχεια, αναζητά μια βελτιωμένη πολιτική έτσι, ώστε να οδηγήσει σε μια βελτιωμένη value function. Αυτή η αλληλουχία policy evaluation – policy improvement επαναλαμβάνεται μέχρι να βρεθεί η βέλτιστη πολιτική και κατ' επέκταση η βέλτιστη value function. Από την άλλη, ο αλγόριθμος Value Iteration ξεκινά με μια τυχαία value function και σε κάθε του επανάληψη

προσπαθεί να βελτιώσει τη συνάρτηση αυτή, χωρίς να ασχολείται με την πολιτική που οδηγεί στις επί μέρους value functions. Φυσικά, δεδομένης της βέλτιστης value function (μόλις συγκλίνει ο αλγόριθμος), ο προσδιορισμός της αντίστοιχης βέλτιστης πολιτικής είναι εύκολος.

Θέμα 3

(α) Ο Hill Climbing αποτελεί έναν ευριστικό, επαναληπτικό αλγόριθμο βελτιστοποίησης. Ξεκινώντας από μια τυχαία λύση ενός προβλήματος, ο αλγόριθμος πραγματοποιεί σταδιακά μικρές διαταραχές στη λύση αυτή, προσπαθώντας σε κάθε βήμα να τη βελτιώσει ως προς κάποιο μέτρο. Η σύγκλιση επέρχεται όταν πλέον οποιαδήποτε μικρή διαταραχή δε δύναται να βελτιστοποιήσει περισσότερο τη λύση. Δεδομένου αυτού, γίνεται εύκολα αντιληπτό πως ο Hill Climbing μπορεί πολύ εύκολα να συγκλίνει σε κάποιο τοπικό ακρότατο, αποτυγχάνοντας να βρει το ολικό βέλτιστο. Το σημείο στο οποίο ο αλγόριθμος Simulated Annealing διαφοροποιείται από τον Hill Climbing είναι πως δύναται να δεχτεί από ένα βήμα για το επόμενο μια λύση η οποία να μην αποτελεί βελτίωση της προηγούμενης βάσει του επιλεγμένου μέτρου. Η διαδικασία αυτή πραγματοποιείται πιθανοκρατικά, με πιθανότητα η οποία εξαρτάται από μια παράμετρο «θερμοκρασίας» η οποία μειώνεται όσο αυξάνεται το πλήθος επαναλήψεων. Στόχος της είναι να θυσιάσει μια βραχυπρόθεσμη βελτίωση, προκειμένου να απομακρυνθεί έτσι από κάποιο τοπικό ακρότατο και τελικά να συγκλίνει στο επιθυμητό ολικό ακρότατο.

Θέμα 4

(α) Η πολύ σημαντική παραδοχή που υιοθετεί ο ταξινομητής Naïve Bayes είναι πως τα χαρακτηριστικά (features) είναι μεταξύ τους υπό συνθήκη ανεξάρτητα. (βλ. Lab 10 για λεπτομέρειες).