Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Κυρτή Βελτιστοποίηση με Εφαρμογές στη Μηχανική Μάθηση
Διδάσκοντες: Κ. Χρυσαφίνος, Δ. Φωτάκης

2η Σειρά Ασκήσεων

Άσκηση 1. (α) Να υπολογίσετε το VC dimension των παρακάτω κλάσεων συναρτήσεων:

- $\mathcal{H}_1 = \{f_{\theta_1,\theta_2} : [0,1] \to \{0,1\} \,|\, 0 < \theta_1 < \theta_2 < 1\}$, όπου $f_{\theta_1,\theta_2}(x) = \begin{cases} 1 & \text{αν } \theta_1 \leq x \leq \theta_2 \\ 0 & \text{διαφορετικά} \end{cases}$
- $\mathcal{H}_2 = \{f_{\theta_1,\theta_2,\theta_3,\theta_4} : [0,1] \to \{0,1\} \,|\, 0 < \theta_1 < \theta_2 < \theta_3 < \theta_4 < 1\}$,
  όπου $f_{\theta_1,\theta_2,\theta_3,\theta_4}(x) = \begin{cases} 1 & \text{αν } \theta_1 \leq x \leq \theta_2 \text{ ή } \theta_3 \leq x \leq \theta_4 \\ 0 & \text{διαφορετικά} \end{cases}$
- $\mathcal{H}_3 = \{f_{\theta_1,\theta_2} : \mathbb{R}^2 \to \{0,1\} \,|\, 0 < \theta_1 < \theta_2\}$, όπου $f_{\theta_1,\theta_2}(x,y) = \begin{cases} 1 & \text{αν } \theta_1 x \leq y \leq \theta_2 x \\ 0 & \text{διαφορετικά} \end{cases}$
- $\mathcal{H}_4$ που περιλαμβάνει όλες τις συναρτήσεις $f : \mathbb{R}^3 \to \{0,1\}$ που κατηγοριοποιούν με βάση κάποιο ορθογώνιο παραλληλεπίπεδο παράλληλο προς τους άξονες (θεωρήστε ότι όσα σημεία βρίσκονται εντός του παραλληλεπιπέδου αποτιμώνται στο 1 και όσα σημεία είναι εκτός αποτιμώνται στο 0).
- $\mathcal{H}_5 = \{f_k : \mathcal{X} \to \{0,1\} \,|\, |\{x \in \mathcal{X} : f_k(x) = 1\}| \leq k\}$. Εξαρτάται το VC dimension της $\mathcal{H}_5$ από τον πληθικό αριθμό του $\mathcal{X}$, και αν ναι, πώς;

(β) Να δείξεται ότι για κάθε κλάση υποθέσεων $\mathcal{H}$, VCdimension$(\mathcal{H} \cup \{f\}) \leq$ VCdimension$(\mathcal{H}) + 1$.

Άσκηση 2. Να αποδείξετε ένα (όσο το δυνατόν μικρότερο) άνω φράγμα για το regret του αλγόριθμου Online Gradient Descent (όπως αυτός παρουσιάζεται στις διαφάνειες της 4ης διάλεξης) λαμβάνοντας υπόψη την προβολή στο κυρτό σύνολο των εφικτών λύσεων S και χρησιμοποιώντας χρονικά μεταβαλλόμενο βήμα $\eta_t = \frac{B}{G\sqrt{t}}$.

Άσκηση 3. Στην μέθοδο Support Vector Machines (SVM), χρησιμοποιούμε ως συνάρτηση κόστους (loss function) την $\ell_{(\vec{x},y)}(\vec{w}) = \max\{0, 1 - y\,\vec{w}\cdot\vec{x}\}$ και υπολογίζουμε τον classifier $\vec{w}$ που ελαχιστοποιεί την παρακάτω συνάρτηση regularized loss:

$$f_{(\vec{x}_1,y_1),\ldots,(\vec{x}_n,y_n)}(\vec{w}) = \frac{\lambda}{n}\sum_{i=1}^{n} \ell_{(\vec{x}_i,y_i)}(\vec{w}) + \frac{\|\vec{w}\|^2}{2}\,,$$

όπου λ είναι το regularization factor και $(\vec{x}_1, y_1), \ldots, (\vec{x}_n, y_n)$ τα δείγματα. Να εξειδικεύσετε τον αλγόριθμο Stochastic Gradient Descent για την μέθοδο Support Vector Machines. Τι βήμα η (ή η_t) θα χρησιμοποιήσετε και τι regret θα επιτύχετε με αυτό;

Άσκηση 4. Θεωρούμε τον παρακάτω αλγόριθμο A για Online Linear Optimization, για τον οποίο θα αναλύσουμε το regret που επιτυγχάνει.

- **Input**: n actions $\{1, \ldots, n\}$, time horizon T, $\vec{w}_1 = (1, \ldots, 1)$, $\vec{x}_1 = (1/n, \ldots, 1/n)$
- for $t = 1$ to T do:
  - Select action $i_t \in \{1, \ldots, n\}$ with probability $\vec{x}_t(i_t)$
  - Get loss $\vec{\ell}_t \in [0,1]^n$ for all actions and incur loss $\vec{\ell}_t(i_t)$
  - Update weights $\vec{w}_{t+1}(i) = \vec{w}_t(i) e^{-\eta \vec{\ell}_t(i)}$, for all $i \in \{1, \ldots, n\}$
  - Update probabilities $\vec{x}_{t+1}(i) = \frac{\vec{w}_{t+1}(i)}{\sum_{i=1}^n w_{t+1}(i)}$, for all $i \in \{1, \ldots, n\}$

(α) Θεωρούμε τη συνάρτηση δυναμικού $\Phi(t) = \sum_{i=1}^n w_t(i)$. Αρχικά είναι $\Phi(1) = n$. Να δείξετε ότι $\Phi(T) \geq e^{-\eta \sum_{t=1}^T \vec{\ell}_t(i^*)}$, όπου $i^* = \arg\min_i \sum_{t=1}^T \vec{\ell}_t(i)$ η βέλτιστη επιλογή.

(β) Να δείξετε ακόμη ότι για κάθε $t \geq 1$,

$$\Phi(t+1) \leq \Phi(t) e^{-\eta \vec{x}_t \cdot \vec{\ell}_t + \eta^2 \vec{x}_t \vec{\ell}_t^2} ,$$

όπου $\vec{\ell}_t^2(i) = (\vec{\ell}_t(i))^2$, για όλα τα $i \in \{1, \ldots, n\}$. Χρησιμοποιώντας αυτή τη σχέση, να βρείτε ένα άνω φράγμα για το $\Phi(T)$ ως συνάρτηση του $\Phi(1) = n$.

(γ) Χρησιμοποιώντας τα άνω και κάτω φράγματα για την $\Phi(T)$ που υπολογίσατε στα (α) και (β) και το γεγονός ότι $\vec{\ell}_t \in [0,1]^n$, για κάθε $t \in \{1, \ldots, T\}$, να δείξετε ότι:

$$\text{Regret}_A(T) \leq \eta T + \ln(n)/\eta$$

Ποια τιμή του η θα επιλέγατε και ποιο είναι το $\text{Regret}_A(T)$ για αυτή;

Υποβολή. Θα πρέπει να έχετε μαζί σας τις απαντήσεις σας στην **εξέταση του μαθήματος** και να τις παραδώσετε εκεί.