



---

## Zadaci

---

### Zadatak 1.

Pokažite ekvivalenciju kriterijskih funkcija

$$J_1(\theta) = \frac{1}{m} \sum_{i=1}^m \log(1 + e^{-y^{(i)}\theta^T x^{(i)}}), \text{ gdje je } y^{(i)} \in \{-1, 1\}$$

i

$$J_2(\theta) = -\frac{1}{m} \sum_{i=1}^m (y^{(i)} \log(h_\theta(x^{(i)})) + (1 - y^{(i)}) \log(1 - h_\theta(x^{(i)}))), \text{ gdje je } y^{(i)} \in \{0, 1\}.$$

### Zadatak 2.

Neka su zadani podaci  $X \in \mathbb{R}^{m \times n}$  i pripadne izlazne vrijednosti  $y \in \mathbb{R}^{m \times k}$  koje označavaju pripadnost nekog podatka jednoj od klasa  $0, 1, \dots, k$ . Na ovim podacima možemo provesti proces učenja klasifikacijskog modela. Jedan od takvih modela je logistička regresija. Logistička regresija kao model funkciju koristi  $h_\theta(x) = \frac{1}{1 + e^{-\theta^T x}}$ , gdje su  $\theta$  parametri koje model uči. Znamo kako se učenje svodi na minimizaciju funkcije  $J(\theta) = \frac{1}{m} \sum_{i=1}^m [-y^{(i)} \log(h_\theta(x^{(i)})) - (1 - y^{(i)}) \log(1 - h_\theta(x^{(i)}))]$  te kako do rješenja tog minimizacijskog problema možemo doći gradijentnom metodom.

1. Implementirajte gradijentnu metodu za slučaj binarne klasifikacije ( $k = 2$ ). Neka vaša funkcija koja pokreće gradijentnu metodu kao argumente prima stopu učenja  $\alpha$  i broj iteracija metode.
2. Isprobajte vašu implementaciju na danom skupu podataka ..

---

U datoteci *heart\_data.csv* nalaze se podaci o pacijentima vezani uz zdravlje njihovog krvožilnog sustava i srca. Svakom pacijentu su navedene godine, spol i podaci koji dolaze iz mjerenja poput razine kolesterola, maksimalni otkucaj srca i slično. Osim toga svakom je pacijentu pridružena oznaka 0 ako ne boluje od srčanih bolesti tj. zdrav je i 1 ako ima srčanu bolest. - Istražite i pripremite ih za učenje modela. (Učitajte, skalirajte, vizualizirajte) - Postavite model logističke regresije koristeći implementaciju iz prethodnog zadatka. - Model je potrebno učiti na 70% podataka, a zatim ga testirati na preostalih 30% podataka. - Kako biste ocijenili model koristite metrike koje smo upoznali na vježbama.

### Zadatak 3.



Na predavanju je predstavljena softmax regresija i dana je kriterijska funkcija

$$J(\theta) = - \sum_{i=1}^m \sum_{j=1}^k I^j(y^{(i)}) \log \frac{e^{\theta^{(j)T} x^{(i)}}}{\sum_{l=1}^k e^{\theta^{(l)T} x^{(i)}}}$$

Raspišite gradijent funkcije  $J(\theta)$ . —

#### Zadatak 4.

a) Neka su zadani podaci  $X \in \mathbb{R}^{m \times n}$  i pripadne izlazne vrijednosti  $y \in \mathbb{R}^{m \times k}$  koje označavaju pripadnost nekog podatka jednoj od klasa  $0, 1, \dots, k$ . Na ovim podacima možemo provesti proces učenja klasifikacijskog modela. Budući da se radi o višeklasnoj klasifikaciji, koristit ćemo *softmax* funkciju. Kao i do sada, učenje se svodi na minimizaciju funkcije

$$J(\theta) = - \sum_{i=1}^m \sum_{j=1}^k I^j(y^{(i)}) \log \frac{e^{\theta^{(j)T} x^{(i)}}}{\sum_{l=1}^k e^{\theta^{(l)T} x^{(i)}}}$$

te kako do rješenja tog minimizacijskog problema možemo doći gradijentnom metodom.

1. Implementirajte gradijentnu metodu za općeniti slučaj klasifikacije. Neka vaša funkcija koja pokreće gradijentnu metodu kao argumente prima stopu učenja  $\alpha$  i broj iteracija metode.
2. Isprobajte implementiranu funkciju na podacima *data\_examples.csv*. Odaberite proizvoljno stopu učenja  $\alpha$  i broj iteracija.
3. Nacrtajte graf promjene vrijednosti funkcije troška kroz iteracije i ispišite vrijednost funkcije troška u zadnjoj iteraciji.

---

b) Zadani su vam podaci koji se nalaze u datoteci *mobiteli*

Ovaj skup sadrži podatke o mobilima - njihove karakteristike i cjenovne rangove. Svaki uzorak (mobil) je opisan kroz 20 različitih varijabli. Varijable su kategoričke poput *dual\_sim* koja poprima vrijednosti 0 i 1 ili numeričke poput *clock\_speed*. 18 varijabli je tipa **int**, a 2 su tipa **float**.

Svaki mobil je klasificiran u jedan cjenovni razred koji može biti 1, 2 ili 3.

Potrebno je upoznati se sa skupom podataka, analizirati ga koristeći neke kvantitativne ili grafičke tehnike. Glavni zadatak je vidjeti može li se na temelju različitih karakteristika mobil odrediti njegov cjenovni razred.

Koristeći *softmax* implementaciju iz a) podzadatka riješite ovaj problem tako da ćete 1. Podatke istražiti i pripremiti ih za učenje modela. (Učitajte, skalirajte, vizualizirajte) 1. Na 80% podataka naučiti model 2. Na 20% podataka testirati model i ispisati matricu zabune, točnost (accuracy)