## Linearna regresija

### Način implementacije algoritma

Klasa "MyLinearRegression" implementira algoritam linearne regresije kroz metode koje obuhvataju inicijalizaciju, prilagođavanje modela, predviđanje i vizualizaciju gubitka.

Klasa sadrži sledeće atribute:

* stopa učenja
* broj iteracija
* tolerancija

\_mean\_squared\_error: Ova pomoćna metoda izračunava srednji kvadratni gubitak (MSE - Mean Squared Error)

fit: Ova metoda koristi gradijentni spust da bi pronašla optimalne koeficijente modela koji minimizuju srednji kvadratni gubitak.

plot\_loss\_history: Ova metoda prikazuje grafikon gubitka po iteraciji, što omogućava vizuelnu proveru konvergencije algoritma.

predict: Nakon što se model prilagodi, ova metoda se koristi za generisanje predviđanja na osnovu ulaznih podataka.

gradient\_descent: Ova metoda implementira algoritam gradijentnog spusta, gde se koeficijenti modela iterativno ažuriraju kako bi se minimizovao gubitak. Algoritam se zaustavlja kada promena gubitka postane manja od zadate tolerancije ili kada se dostigne maksimalan broj iteracija.

### Skup podataka

Ovaj skup podataka se može koristiti da se predviđaju troškovi osiguranja. Skup podataka ima 1338 uzoraka i 7 atributa. Skup podataka ima sledeće atribute:

* age: starost primarnog korisnika
* pol: Pol ugovarača osiguranja (muški ili ženski)
* bmi: Indeks telesne mase
* children: broj dece obuhvaćene zdravstvenim osiguranjem
* smoker: Da li je osiguranih pušač ili ne
* region: stambeno područje korisnika u SAD-u (severoistok, jugoistok, jugozapad, severozapad)
* troškovi: Individualni medicinski troškovi koje fakturiše zdravstveno osiguranje

Kako nisu svi podaci numeričkog tipa, u procesu pretprocesiranja podataka izvršeno je enkodiranje nenumeričkih feature-a.

### Zaključak

Na Figure 1 su prikazane vrednosti za srednju kvadratnu grešku za moju implementaciju algoritma i za korišćenu gotovu funkciju iz biblioteke scikit learn.



Figure 1

Na Figure 2 je prikazana vrednost srednje kvadratne greške po iteraciji. Na osnovu grafika se može zaključiti da veličina greške opada sa povećanjem broja iteracija, ali da posle 1000 iteracije dolazi do zaravnjenja krive, što može značiti da dolazi do overfitting-a modela.

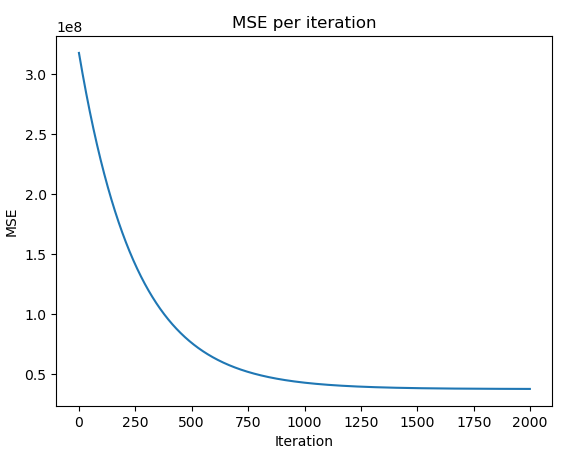


Figure 2