**Priprema podataka za obučavanje**

***get\_table\_of\_location\_distance\_from\_center( )*** - fuckija provera da li smo kreirali tabelu u bazi koja ima podatke o ulicama I lokacijama u Beogradu, I njihove udaljenosti od centra. Ukoliko tabela ne postoji ili se obrisala, čitamo iz baze sve različite lokacije u Beogradu, preko sajta Google Maps dohvatamo udaljenosti i upisujemo u bazu. Ovo je neophodno da bi dohvatanje udaljenosti lokacije od centra bilo efikasnije.

***training\_data()*** – funkcija prolazi kroz 70% svih redova u bazi koji predstavljaju prodaju stanova na teritoriji Beograda. Zahvaljujući prethodnoj funkciji dohvata se i udaljenosti lokacija od centra. Dalje se vrši uniformisanje podataka, za svaku odliku se vrši ažuriranje minimalnih i maksimalnih vrednosti da bi se podaci normalizovali funkcijom **normalize()** pre nego što se izvrši linearna regresija. Relevantni podaci iz baze koji prestavlaju odlike (površina, stanje objekta, godina izgradnje, sprat, ukupna spratnost, uknjiženost, broj soba, parking, lift, udaljenost lokacije) se stavljaju u matricu x, dok se cene stavljaju u niz y.

**Realizacija višestruke linearne regresije i pomoćne funkcija**

U literaturi odlike modela se predstavljaju kao niz . U programu su odlike predstavljenje kao niz [], gde imamo 10 članova, jer imamo 10 odlika na osnovu kojih se traži hipoteza linearne regresije.

***h(x, i)*** –pomoćna funkcija koja računa hipotezu višestruke univarijantne linearne regreseije za trenutno aktuelne vrednosti *w*[].

***j\_mse( )*** – funkcija koja računa funkciju greške kao prosek kvadrata odstupanja *h(x)* od *y* (eng. *Mean Sqared Error - MSE*): po formuli .

***gradient\_descent( )*** – funkcija gradijentnog spusta koja predstavlja iterativni metod traženja minumina funkcije. Na samom početku inicijalizujemo nizove *w*[] *temp*[] vrednostima nula. Ažuriramo te vrednosti sledećim postupkom sve dok postoji opadanje funkcije greške (koju računa fukncija ***j\_mse( )***), tj. dok je vrednost trenutne funkcije greške umanjena funckijom greške za prethodne paramtre manja od 0. Postupak ažuriranja vrenodosti parametara *temp* vrše se prema sledećim formulama:

Pomoćna funkcija ***sum\_h\_minus\_y()*** računa sume , …, .

Kada se preračunaju temp vrednosti, onda se ažuriraju i w vrednosti:

Potom se računa funkcija greške za ažurirane vrednosti w, i ukoliko je ona manje od funkcije greške za prethodne vrednosti parametara w, ide se u smeru konvergencije i nastavlja se proces.

**Testiranje**

***test\_data()*** – za preostalih 30% podataka u bazi koji predstavljaju prodaju stanova na teritoriji Beograda radi se uniformisanje podataka, normalizacija i računanje koja bi bila predviđena cena, kao i greška nastala tom prilikom.

**Forma za unos podataka**

***input\_form()*** – postavljaju se pitanja korisniku da bi uneo tražene parametre koji se normalizuju, i za njih se izračunava dobijena hipoteza višestruke univarijantne linearne regresije. Rezultat se prikazuje korisniku kao predviđena cena nekretnine.

Linearna regresija je u ovom zadatku proučavana:

1. Sa jedom odlikom (kvadratura)
2. Sa 2 odlike (kvadratura, lokacija)
3. Sa 3 odlike (kvadratura, broj soba, lokacija)
4. Sa 10 odlika (kvadratura, stanje objekta, godina izgradnje, sprat, ukupna spratnost, uknjiženost, broj soba, parking, lift, lokacija).