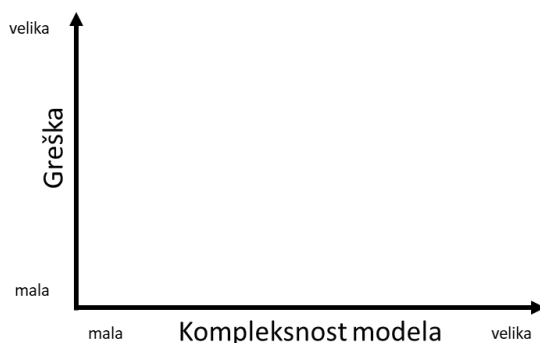


- Tačno ili netačno:
 - Overfitting* je verovatniji kada postoji velika količina podataka za treniranje.
 - Overfitting* je verovatniji ako imamo veliki broj obeležja.
 - Overfitting* je verovatniji ako koristimo fleksibilan (kompleksan) model.
 - Overfitting* je verovatniji ako smo uočili da su vrednosti parametara dobijenog modela male.
 - Verovatnije je da će model $y = \theta_0$ patiti od velike varijanse nego od velikog sistematskog odstupanja.
 - Za fiksni model, ako povećavamo trening skup, očekujemo da vrednosti parametara modela θ rastu.
- Zamislite da ste na skupu podataka trenirali model linearne regresije 3. stepena polinoma i otkrili da su trening i test greška jednake 0 (odnosno, da se model savršeno uklapa u podatke).
 - Šta očekujete za polinom 4. stepena? (1) verovatnoća *overfitting*-a je velika; (2) verovatnoća *underfitting*-a je velika; (3) ne možemo reći; (4) nijedan od ponuđenih odgovora.
 - Šta očekujete za polinom 2. stepena? (1) verovatnoća *overfitting*-a je velika; (2) verovatnoća *underfitting*-a je velika; (3) ne možemo reći; (4) nijedan od ponuđenih odgovora.
 - Šta će se desiti ako koristite polinom 2. stepena? Sistematsko odstupanje će biti _____, varijansa će biti _____. (malo/veliko)
- Na grafiku levo nacrtajte kako se ponaša (a) trening (b) test greška. Na grafiku desno nacrtajte kako se ponaša (a) trening i (b) test greška ako je **kompleksnost modela fiksirana**.



- Na levom grafiku označite koji deo odgovara *overfitting*-u, a koji *underfitting*-u.
- Na levom grafiku označite koji deo odgovara velikom sistematskom odstupanju, a koji varijansi.
- Recimo da ste odabrali model optimalne kompleksnosti. Nakon toga ste identifikovali obeležje koje je veoma relevantno za rešavani zadatak. Šta očekujete da će se desiti ako dodate ovo obeležje u model? Trening greška će se _____. Greška na validacionom skupu će se _____. (uvećati/smanjiti)
- Za svaku od tehnika naznačite da li imaju potencijal da **redukuju** sistematsko odstupanje (bias) i/ili varijansu:

	Bias	Varijansa
Uvećanje fleksibilnosti modela (npr. dodavanje slojeva/neurona u neuronsku mrežu)		
Primena <i>feature engineering</i> (npr. izmena ili dodavanje obeležja prema uvidima proisteklim iz analize grešaka modela)		
Selekcija obeležja da bi se smanjio broj obeležja		
<i>Early stopping</i> (zaustaviti <i>gradient descent</i> ranije, zasnovano na grešci validacionog skupa)		
Izmena modela (npr. izmena arhitekture neuronske mreže)		
Dodavanje više trening podataka		

8. Možemo li **linearnom** regresijom modelovati **kvadratnu** funkciju?
9. Trening skup služi za _____, validacioni skup služi za _____, a test skup služi za _____. Test skupom aproksimiramo _____ grešku.
10. Šta je „idealna“ stratifikacija, a šta obično radimo u praksi?
11. Kako biste stratifikovali podatke za regresiju, a kako za klasifikaciju?
12. Kada biste koristili podelu na trening i test skup, a kada unakrsnu validaciju?
13. Imamo skup podataka. Radi brže konvergencije *gradient descent*-a normalizovali smo vrednosti obeležja. Nakon toga smo podelili skup podataka na trening, validacioni i test skup. Da li smo sve dobro uradili?
14. U slučaju **višestruke** regresije, šta je ključna stvar o kojoj moramo voditi računa prilikom interpretacije parametara modela?
15. Kako enkodiramo nominalna, a kako ordinalna kategorička obeležja?
.
16. Šta je potencijalni problem prethodnog pristupa? Kako biste ga rešili?
17. Multikolinearnost:
 - a. Nije problematična.
 - b. Nije problematična za performanse modela, ali troši računarske resurse.
 - c. Je nezgodna pri interpretaciji modela.
 - d. Može da se desi kod 1-hot-encoding-a.