

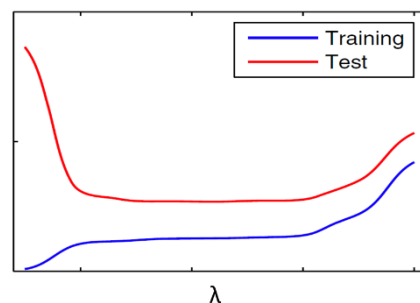
1. Navedite barem tri mere kvaliteta modela koje biste koristili u slučaju klasifikacije:

2. Navedite barem dve mere kvaliteta modela koje biste koristili u slučaju regresije:

3. Popunite tabelu (za definicije ne morate stavljati formule, u redu je da opišete smisao rečima)

| Metrika    | Definicija | Primer realne primene gde biste datu metriku upotrebili kao najbolju za evaluaciju modela |
|------------|------------|---|
| Preciznost |            |   |
| Odziv      |            |   |
| F-mera     |            |   |
| RMSE       |            |   |
| $R^2$      |            |   |

4. Grafik prikazuje trening i test grešku za model linearne regresije sa regularizacionim parametrom  $\lambda$ :  $J(\theta) = \frac{1}{2N} \left( \sum_{i=1}^N (h_{\theta}(x^{(i)}) - y^{(i)})^2 + \lambda \sum_{j=1}^d \theta_j^2 \right)$ . Označite za koju vrednost  $\lambda$  model pati od: (1) velikog sistematskog odstupanja (*bias*), (2) velike varijanse (*variance*).



5. Opišite kako dobijamo *learning curves* i kako ih koristimo za ML dijagnostiku (skicirajte primer za veliku varijansu i primer za veliko sistematsko odstupanje).

6. Projektni tim je izvršio selekciju važnih obeležja nad celim skupom podataka i značajno smanjio broj obeležja. Nakon toga, podelili su podatke u trening i test skup. Zatim su optimizovali više modela na trening skupu i evaluirali njihove performanse na test skupu. Kao performanse modela prijavljene su performanse najboljeg modela postignute nad test skupom. Da li je ova postavka korektna? Obrazložite.

7. Koje od navedenih strategija su adekvatne strategije za rešavanje problema velikog sistematskog odstupanja, a koje za rešavanje problema velike varijanse:

|   | 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   |      |           |
| Smanjenje kompleksnosti modela  |      |           |
| Smanjenje šuma u trening skupu  |      |           |
| Smanjenje ili uklanjanje regularizacije   |      |           |
| Dodavanje regularizacije  |      |           |

8. Šta označava pojam Okamova oštrica u kontekstu mašinskog učenja?

9. Šta označava pojam *sampling bias*?

10. Navedite primer *sampling bias*-a.

11. Kako rešavamo problem *sampling bias*-a?

12. Šta označava pojam *data snooping*?

13. Kako rešavamo problem *data snooping*-a?