

## Rešenje zadatka - vežbe 5, 2021.

**Zadatak:** Proceniti performanse dva klasifikaciona modela,  $M_1$  i  $M_2$ . Skup podataka sadrži 26 binarnih atributa, označenih od A do Z. Tabela prikazuje posteriorne verovatnoće za pozitivnu klasu dobijene primenom modela na skup podataka. Nacrtati ROC krive.

Instance	True Class	$P(+ A, \dots, Z, M_1)$	$P(+ A, \dots, Z, M_2)$
1	+	0.73	0.61
2	+	0.69	0.03
3	-	0.44	0.68
4	-	0.55	0.31
5	+	0.67	0.45
6	+	0.47	0.09
7	-	0.08	0.38
8	-	0.15	0.05
9	+	0.45	0.01
10	-	0.35	0.04

### Mere koje se koriste za crtanje ROC krive

- TP - true positive (broj instanci koje pripadaju pozitivnoj klasi i dodeljena im je pozitivna klasa)
- FP - false positive (broj instanci koje pripadaju negativnoj klasi, a dodeljena im je pozitivna klasa)
- TN - true negative (broj instanci koje pripadaju negativnoj klasi i dodeljena im je negativna klasa)
- FN - false negative (broj instanci koje pripadaju pozitivnoj klasi, a dodeljena im je negativna klasa)

Matrica konfuzije za binarnu klasifikaciju (pozitivna i negativna klasa)

		Dodeljena klasa	
		+	-
Prava klasa	+	TP	FN
	-	FP	TN

$$TPR = \frac{TP}{TP+FN} \text{ (stopa stvarno pozitivnih)}$$

$$FPR = \frac{FP}{FP+TN} \text{ (stopa lažno pozitivnih)}$$

## Crtanje ROC krive za jedan model

Da bi mogla da se nacrtaja ROC kriva, potrebno je da model za klasifikaciju za instancu koju klasifikuje daje i neprekidnu vrednost kao izlaz koji se može koristiti za uređivanje instanci prema predviđanju pripadnosti instance klasi, npr. verovatnoću da instanca pripada određenoj klasi. Procedura za crtanje ROC krive je:

- Test instance se uredi u rastućem poretku na osnovu neprekidnog izlaza (u zadatku je to verovatnoća da instanca pripada pozitivnoj klasi, pa će se na dalje umesto dodeljenog neprekidnog izlaza instanci navoditi dodeljena verovatnoća ) koji im je dodelio model.
- Nakon uređenja po verovatnoći, u svakom koraku se na sledeći način obrađuje po jedna test instanca, počevši od instance sa najmanjom dodeljenom verovatnoćom:
  - Instanci koja se obrađuje, kao i svakoj instanci sa većom verovatnoćom (tj. svim instancama u tabeli udesno od instance koja se obrađuje) se dodeljuje pozitivna klasa, a svakoj instanci sa manjom verovatnoćom (tj. svim instancama u tabeli ulevo od instance koja se obrađuje) se dodeljuje negativna klasa.
  - Sa tako dodeljenim klasama i podacima o pravim klasama, računaju se TPR i FPR.
- Dobijene vrednosti (TPR i FPR) u svakom koraku se predstavljaju kao posebna tačka na grafiku. ROC kriva se dobija spajanjem tačaka koje predstavljaju rezultat obrade dve susedne test instance.

### Crtanje ROC krive za model $M_1$ :

Korak 1: Obraduje se 1. instanca, tako da se njoj i svim instancama desno od nje (tj. svim instancama) dodeljuje pozitivna klasa.

Verovatnoća	<b>0,08</b>	0,15	0,35	0,44	0,45	0,47	0,55	0,67	0,69	0,73	
Prava klasa	-	-	-	-	+	+	-	+	+	+	
Dodeljena klasa	+	+	+	+	+	+	+	+	+	+	
TP	5										
FP	5										
TN	0										
FN	0										
TPR	<b>1</b>										
FPR	<b>1</b>										

Korak 2: Obraduje se 2. instanca, tako da se njoj i svim instancama desno od nje dodeljuje pozitivna klasa, a instanci levo od nje (1. instanci) negativna klasa.

Verovatnoća	0,08	<b>0,15</b>	0,35	0,44	0,45	0,47	0,55	0,67	0,69	0,73	
Prava klasa	-	-	-	-	+	+	-	+	+	+	
Dodeljena klasa	-	+	+	+	+	+	+	+	+	+	
TP	5	5									
FP	5	4									
TN	0	1									
FN	0	0									
TPR	1	<b>1</b>									
FPR	1	<b>0,8</b>									

Korak 3: Obraduje se 3. instanca, tako da se njoj i svim instancama desno od nje dodeljuje pozitivna klasa, a instancama levo od nje (1. i 2. instanci) negativna klasa.

Verovatnoća	0,08	0,15	<b>0,35</b>	0,44	0,45	0,47	0,55	0,67	0,69	0,73	
Prava klasa	-	-	-	-	+	+	-	+	+	+	
Dodeljena klasa	-	-	+	+	+	+	+	+	+	+	
TP	5	5	5								
FP	5	4	3								
TN	0	1	2								
FN	0	0	0								
TPR	1	1	<b>1</b>								
FPR	1	0,8	<b>0,6</b>								

Korak 4: Obraduje se 4. instanca, tako da se njoj i svim instancama desno od nje dodeljuje pozitivna klasa, a instancama levo od nje (1, 2. i 3. instanci) negativna klasa.

Verovatnoća	0,08	0,15	0,35	<b>0,44</b>	0,45	0,47	0,55	0,67	0,69	0,73	
Prava klasa	-	-	-	-	+	+	-	+	+	+	
Dodeljena klasa	-	-	-	+	+	+	+	+	+	+	
TP	5	5	5	5							
FP	5	4	3	2							
TN	0	1	2	3							
FN	0	0	0	0							
TPR	1	1	1	<b>1</b>							
FPR	1	0,8	0,6	<b>0,4</b>							

Korak 5: Obraduje se 5. instanca, tako da se njoj i svim instancama desno od nje dodeljuje pozitivna klasa, a instancama levo od nje negativna klasa.

Verovatnoća	0,08	0,15	0,35	0,44	<b>0,45</b>	0,47	0,55	0,67	0,69	0,73	
Prava klasa	-	-	-	-	+	+	-	+	+	+	
Dodeljena klasa	-	-	-	-	+	+	+	+	+	+	
TP	5	5	5	5	5						
FP	5	4	3	2	1						
TN	0	1	2	3	4						
FN	0	0	0	0	0						
TPR	1	1	1	1	<b>1</b>						
FPR	1	0,8	0,6	0,4	<b>0,2</b>						

Korak 6: Obraduje se 6. instanca, tako da se njoj i svim instancama desno od nje dodeljuje pozitivna klasa, a instancama levo od nje negativna klasa.

Verovatnoća	0,08	0,15	0,35	0,44	0,45	<b>0,47</b>	0,55	0,67	0,69	0,73	
Prava klasa	-	-	-	-	+	+	-	+	+	+	
Dodeljena klasa	-	-	-	-	-	+	+	+	+	+	
TP	5	5	5	5	5	4					
FP	5	4	3	2	1	1					
TN	0	1	2	3	4	4					
FN	0	0	0	0	0	1					
TPR	1	1	1	1	1	<b>0,8</b>					
FPR	1	0,8	0,6	0,4	0,2	<b>0,2</b>					

Korak 7: Obraduje se 7. instanca, tako da se njoj i svim instancama desno od nje dodeljuje pozitivna klasa, a instancama levo od nje negativna klasa.

Verovatnoća	0,08	0,15	0,35	0,44	0,45	0,47	<b>0,55</b>	0,67	0,69	0,73	
Prava klasa	-	-	-	-	+	+	-	+	+	+	
Dodeljena klasa	-	-	-	-	-	-	+	+	+	+	
TP	5	5	5	5	5	4	3				
FP	5	4	3	2	1	1	1				
TN	0	1	2	3	4	4	4				
FN	0	0	0	0	0	1	2				
TPR	1	1	1	1	1	0,8	<b>0,6</b>				
FPR	1	0,8	0,6	0,4	0,2	0,2	<b>0,2</b>				

Korak 8: Obraduje se 8. instanca, tako da se njoj i svim instancama desno od nje dodeljuje pozitivna klasa, a instancama levo od nje negativna klasa.

Verovatnoća	0,08	0,15	0,35	0,44	0,45	0,47	0,55	<b>0,67</b>	0,69	0,73	
Prava klasa	-	-	-	-	+	+	-	+	+	+	
Dodeljena klasa	-	-	-	-	-	-	-	+	+	+	
TP	5	5	5	5	5	4	3	3			
FP	5	4	3	2	1	1	1	0			
TN	0	1	2	3	4	4	4	5			
FN	0	0	0	0	0	1	2	2			
TPR	1	1	1	1	1	0,8	0,6	<b>0,6</b>			
FPR	1	0,8	0,6	0,4	0,2	0,2	0,2	<b>0</b>			

Korak 9: Obraduje se 9. instanca, tako da se njoj i svim instancama desno od nje dodeljuje pozitivna klasa, a instancama levo od nje negativna klasa.

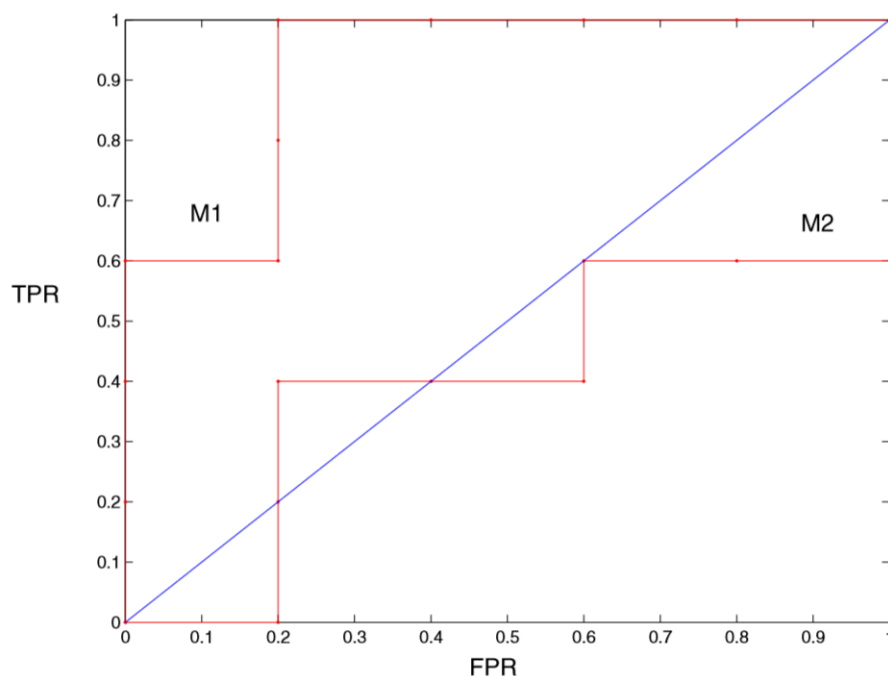
Verovatnoća	0,08	0,15	0,35	0,44	0,45	0,47	0,55	0,67	<b>0,69</b>	0,73	
Prava klasa	-	-	-	-	+	+	-	+	+	+	
Dodeljena klasa	-	-	-	-	-	-	-	-	+	+	
TP	5	5	5	5	5	4	3	3	2		
FP	5	4	3	2	1	1	1	0	0		
TN	0	1	2	3	4	4	4	5	5		
FN	0	0	0	0	0	1	2	2	3		
TPR	1	1	1	1	1	0,8	0,6	0,6	<b>0,4</b>		
FPR	1	0,8	0,6	0,4	0,2	0,2	0,2	0	<b>0</b>		

Korak 10: Obraduje se 10. instanca, tako da se njoj dodeljuje pozitivna klasa, a instancama levo od nje negativna klasa.

Verovatnoća	0,08	0,15	0,35	0,44	0,45	0,47	0,55	0,67	0,69	<b>0,73</b>	
Prava klasa	-	-	-	-	+	+	-	+	+	+	
Dodeljena klasa	-	-	-	-	-	-	-	-	-	+	
TP	5	5	5	5	5	4	3	3	2	1	
FP	5	4	3	2	1	1	1	0	0	0	
TN	0	1	2	3	4	4	4	5	5	5	
FN	0	0	0	0	0	1	2	2	3	4	
TPR	1	1	1	1	1	0,8	0,6	0,6	0,4	<b>0,2</b>	
FPR	1	0,8	0,6	0,4	0,2	0,2	0,2	0	0	<b>0</b>	

Korak 11: Svim instancama se dodeljuje negativna klasa.

Verovatnoća	0,08	0,15	0,35	0,44	0,45	0,47	0,55	0,67	0,69	0,73	
Prava klasa	-	-	-	-	+	+	-	+	+	+	
Dodeljena klasa	-	-	-	-	-	-	-	-	-	-	
TP	5	5	5	5	5	4	3	3	2	1	0
FP	5	4	3	2	1	1	1	0	0	0	0
TN	0	1	2	3	4	4	4	5	5	5	5
FN	0	0	0	0	0	1	2	2	3	4	5
TPR	1	1	1	1	1	0,8	0,6	0,6	0,4	0,2	<b>0</b>
FPR	1	0,8	0,6	0,4	0,2	0,2	0,2	0	0	0	<b>0</b>



Na grafiku je prikazana ROC kriva za model  $M_1$ , kao i za model  $M_2$ . U nastavku je data tabela sa izračunatim TPR i FPR vrednostima za model  $M_2$  koje su potrebne za crtanje ROC krive.

Izračunate potrebne vrednosti za model  $M_2$ :

Verovatnoća	0,01	0,03	0,04	0,05	0,09	0,31	0,38	0,45	0,61	0,68	
Klasa	+	+	-	-	+	-	-	+	+	-	
TP	5	4	3	3	3	2	2	2	1	0	0
FP	5	5	5	4	3	3	2	1	1	1	0
TN	0	0	0	1	2	2	3	4	4	4	5
FN	0	1	2	2	2	3	3	3	4	5	5
TPR	1	0,8	0,6	0,6	0,6	0,4	0,4	0,4	0,2	0	0
FPR	1	1	1	0,8	0,6	0,6	0,4	0,2	0,2	0,2	0

**Zaključak:** Model  $M_1$  je bolji jer je površina ispod ROC krive (AUC) za model  $M_1$  veća nego za model  $M_2$ .