Mere kvaliteta modela

Procena kvaliteta i selekcija modela

- Procena kvaliteta modela se bavi ocenom greške predviđanja modela
- Selekcija modela se bavi izborom jednog od više mogućih modela
- Zasnivaju se na:
  - Merama kvaliteta (npr. tačnost)
  - Tehnikama evaluacije i izbora (npr. podela na trening/test skup)

Važnost numeričke evaluacije

- Idealno: kvalitet modela izražen je jednim realnim brojem

- Recimo da pokušavamo da utvrdimo da li da reči discount/discounts/discounted/discounting tretiramo kao istu reč
  - Sve reči počinju sa istim nizom slova pa nam to može pokazati da su bliske
  - Ali možemo i pogrešiti, npr. universe/university
  - Da li koristiti ovaj prisup ili ne je jako teško reći analizom grešaka modela. Jedini način da to uradimo jeste da isprobamo i vidimo šta bolje radi
  - Za ovo nam treba numerička evaluacija (npr. greška na validacionom skupu)
  - Recimo da sa ovim pristupom dobijemo grešku 3%, a bez njega 5% → izgleda da ovaj pristup jeste dobra ideja

- Takođe, analiza grešaka modela zahteva vreme. Ako isprobavamo mnogo stvari (naročito u početku), lakše je da imamo brz način evaluacije

Mere kvaliteta

- Zavise od problema

- Za klasifikaciju:
  - Tačnost
  - F_1 mera
  - Preciznost/odziv

- Za regresiju:
  - Koren srednjekvadratne greške *RMSE*
  - Koeficijent determinacije R^2

Tačnost

- Nije dobar izbor u slučaju neizbalansiranih (*skewed*) kategorija
- Slučaj gde je jedna (ili više) kategorija u skupu podataka mnogo više zastupljena u odnosu na ostale
- Npr. treba da razvijemo model za detekciju raka
  - Trenirali smo model i na test skupu dobili tačnost od 99%
  - Međutim, recimo da utvrdimo da svega 0.5% pacijenata u našem skupu podataka ima rak
  - greška od 1% više ne deljuje impresivno jer bi model koji sve pacijente klasifikuje u kategoriju $y = 0$ imao 99.5% tačnosti, a ovaj model ne radi ništa korisno!
- Dakle, u slučaju neizbalansiranih klasa, tačnost nije dobra metrika i treba je zameniti nekom drugom

Tabela kontigencije (*contingency table*)

	Tačna klasa	
Predikcija	**1**	0
1	tp	fp
0	fn	tn

- Obično se za $y = 1$ uzima ređa klasa
- **tp** (*true positive*): broj instanci za koje je model predvideo $y = 1$ i stvarna klasa tih instanci je zaista 1
- **fp** (*false positive*): broj instanci za koje je model predvideo $y = 1$, a stvarna klasa tih instanci je 0
- **tn** (*true negative*): broj instanci za koje je model predvideo $y = 0$ i stvarna klasa tih instanci je 0
- **fn** (*false negative*): broj instanci za koje je model predvideo $y = 0$, a stvarna klasa tih instanci je 1

Preciznost (*precision*) i odziv (*recall*)

		Tačna klasa	
Predikcija		**1**	0
	1	$\boxed{\text{tp}}$	fp
	0	fn	tn

Deo pacijenata za koje je $\hat{y} = 1$ i zaista imaju rak $y = 1$

Preciznost (*precision*) i odziv (*recall*)

	Tačna klasa	
Predikcija	**1**	0
1	tp	fp
0	fn	tn

Svi pacijenti za koje je model predvideo $\hat{y} = 1$ (ima rak)

Deo pacijenata za koje je $\hat{y} = 1$ i zaista imaju rak $y = 1$

Preciznost

Da li smo greškom za zdravog pacijenta predvideli da ima rak?

$$precision = \frac{tp}{tp + fp}$$

Preciznost (*precision*) i odziv (*recall*)

Svi pacijenti za koje je
$y = 1$ (zaista imaju rak)

	Tačna klasa	
Predikcija	**1**	0
1	$\boxed{\text{tp}}$	fp
0	fn	tn

Deo pacijenata za koje je $\hat{y} = 1$ i zaista imaju rak $y = 1$

Preciznost

Da li smo greškom za zdravog pacijenta predvideli da ima rak?

Odziv

Da li smo promašili da detektujemo rak kod nekog pacijenta?

$$precision = \frac{tp}{tp + fp}$$

$$recall = \frac{tp}{tp + fn}$$

Dobar model ima i veliku preciznost i veliki odziv

Preciznost (*precision*) i odziv (*recall*)

$$accuracy = \frac{tp + tn}{tp + fp + tn + fn} \qquad precision = \frac{tp}{tp + fp} \qquad recall = \frac{tp}{tp + fn}$$

Imamo 100 pacijenata od kojih 5 zaista ima rak

Model koji predviđa da nijedan pacijent nema rak:

	Tačna klasa		
Predikcija		1	0
	1	0	0
	0	5	95

$$accuracy = 99.5\%$$
$$precison = 0$$
$$recall = 0$$

Model koji predviđa da svi pacijenti imaju rak:

	Tačna klasa		
Predikcija		1	0
	1	5	95
	0	0	0

$$accuracy = 0.5\%$$
$$precison = 0.05$$
$$recall = 1$$

Nagodba preciznosti i odziva

$$\hat{y} = \begin{cases} 1, & h_\theta \geq 0.5 \\ 0, & h_\theta < 0.5 \end{cases}$$

1. Predviđamo $\hat{y} = 1$ samo ako smo sigurni u predikciju

$$\hat{y} = \begin{cases} 1, & h_\theta \geq 0.7 \\ 0, & h_\theta < 0.7 \end{cases}$$

- Izbegavamo fp – naći ćemo malo pacijenata sa rakom, ali smo za te pacijente sigurni da imaju rak

- Veća preciznost
- Manji odziv

2. Izbegavamo da promašujemo previše slučajeva raka

$$\hat{y} = \begin{cases} 1, & h_\theta \geq 0.3 \\ 0, & h_\theta < 0.3 \end{cases}$$

- Izbegavamo *fn* – naći ćemo gotovo sve pacijente sa rakom, ali ćemo za dosta njih koji nemaju rak (pogrešno) predvideti da ga imaju

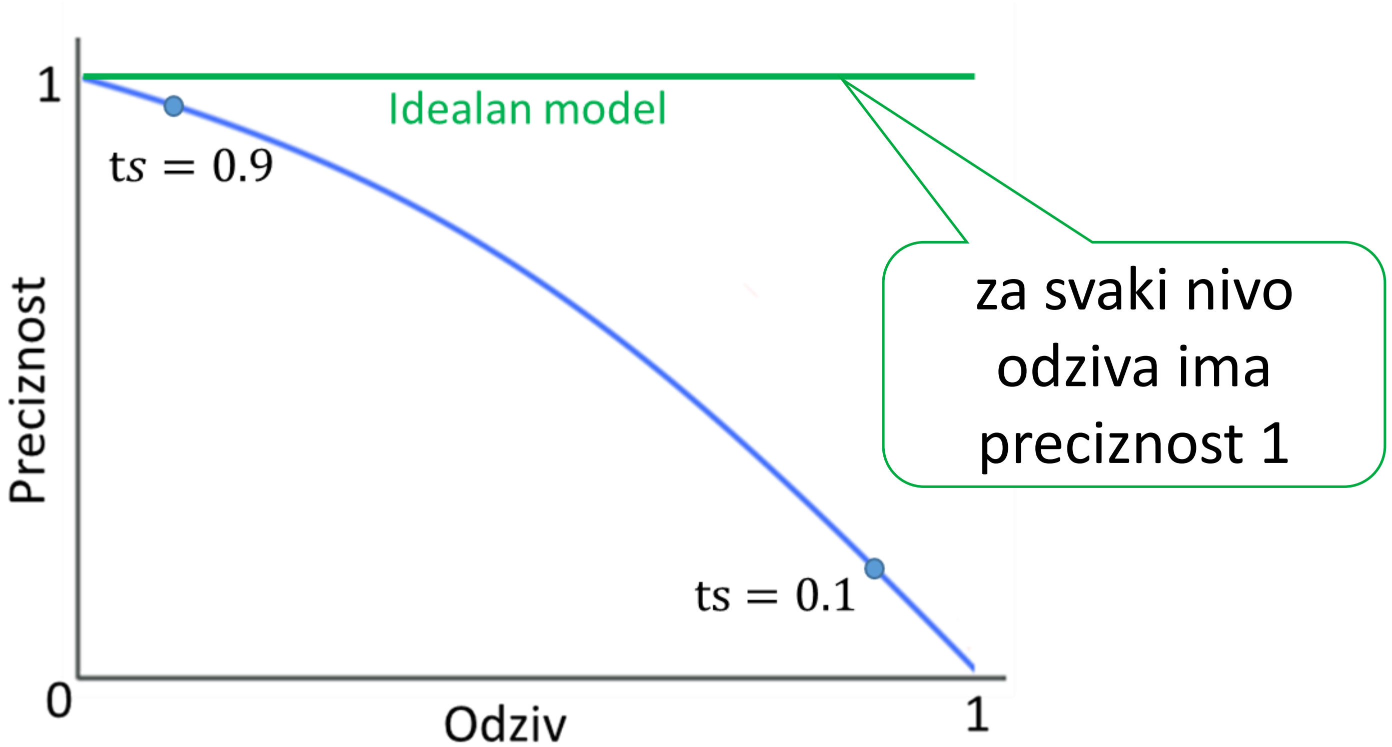
- Manja preciznost
- Veći odziv

Nagodba preciznosti i odziva

- Za jedan model h_θ možemo da balansiramo između preciznosti i odziva:

$$\hat{y} = \begin{cases} 1, & h_\theta \geq \text{ts} \\ 0, & h_\theta < \text{ts} \end{cases}$$

  - Veći *threshold* → veća preciznost. Manji *threshold* → veći odziv

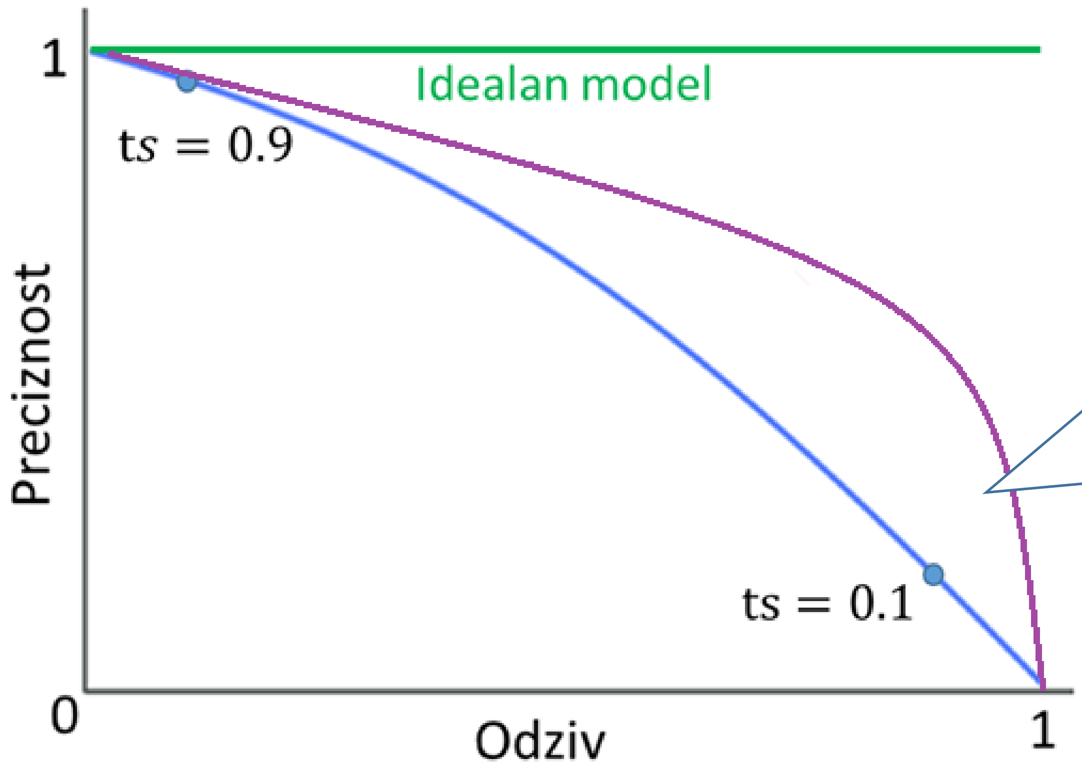


Nagodba preciznosti i odziva

- Za jedan model h_θ možemo da balansiramo između preciznosti i odziva:

$$\hat{y} = \begin{cases} 1, & h_\theta \geq \text{ts} \\ 0, & h_\theta < \text{ts} \end{cases}$$

  - Veći *threshold* → veća preciznost. Manji *threshold* → veći odziv

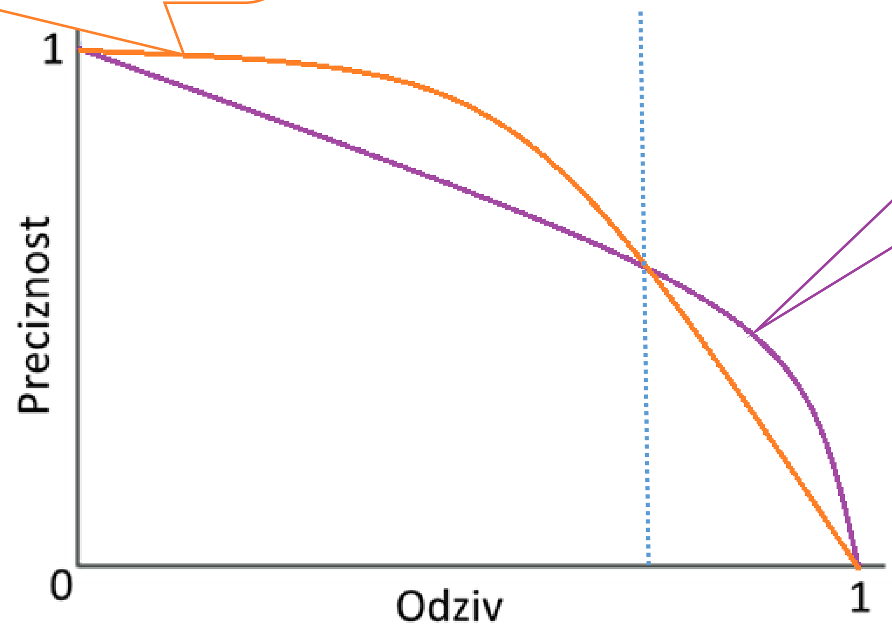


ljubičasti model je bolji od plavog – u svakoj tački daje veću preciznost za isti odziv

Nagodba preciznosti i odziva

Ako nam je bitna visoka preciznost, a ne smeta nam mali odziv, treba da izaberemo narandžasti

Ali ako nam je važan veliki odziv, treba da odaberemo ljubičasti



Za poređenje bi nam bilo lakše da preciznost i odziv sumarizujemo u jedinstven broj

F_1 mera

	Preciznost (P)	**Odziv (R)**	$(P+R)/2$	**F_1 mera**
Model 1	0.5	0.4	0.45	0.444
Model 2	0.7	0.1	0.4	0.175
Model 3	0.02	1.0	0.51	0.0392

Prosek nije najbolje rešenje.

- Ovaj model stalno predviđa $y = 1$ pa imamo jako veliko R i jako malo P
- Model nije dobar a ipak ima najveću vrednost od tri poređena modela

Kombinuje preciznost i odziv, ali je bliže manjoj od te dve mere

$$F_1 score = 2\frac{PR}{P+R}$$

Ako je ili P ili R jednako nuli, i F_1 mera je jednaka nuli

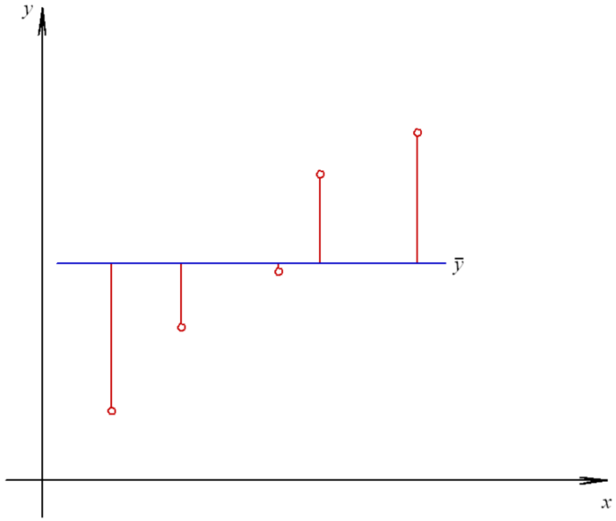
Ako je P =1 i R = 1 (savršen model), onda je F_1 = 1

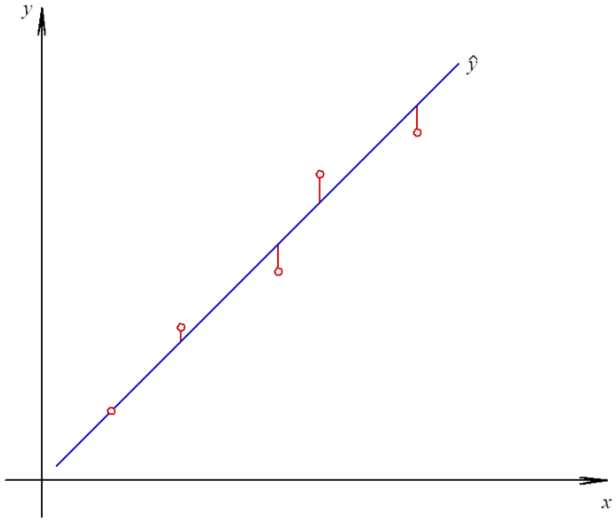
F_1 mera nije osetljiva na neizbalansiranost klasa

Koren srednjekvadratne greške (*RMSE*)

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(h_\theta(x^{(i)}) - y^{(i)})^2}$$

- Poput standardne devijacije, ali ne u odnosu na prosek, već u odnosu na model
- Izražava se u istim jedinicama kao i ciljna promenljiva
- Posebno korisna ako znamo prihvatljivu veličinu greške u razmatranoj primeni





Koeficijent determinacije R^2

$$R^2 = 1 - \frac{MSE}{Var} = 1 - \frac{\sum_{i=1}^{N}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2}{\sum_{i=1}^{N}(\bar{y} - y^{(i)})^2}$$

- U rasponu $(-\infty, 1]$ - meri udeo varijanse ciljne promenljive koji je objašnjen modelom
  - Na primer, ako predviđamo cenu kuće (y) na osnovu kvadrature (x) i dobijemo $R^2 = 0.8$, to znači da je približno 80% varijabilnosti u cenama kuća objašnjeno pomoću informacija o kvadraturi tih kuća
- Pogodnija za poređenja nego kao apsolutna mera