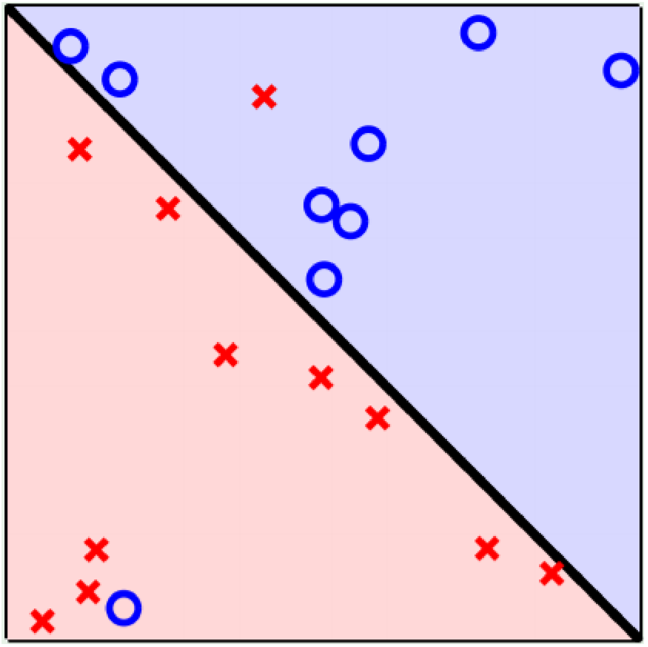
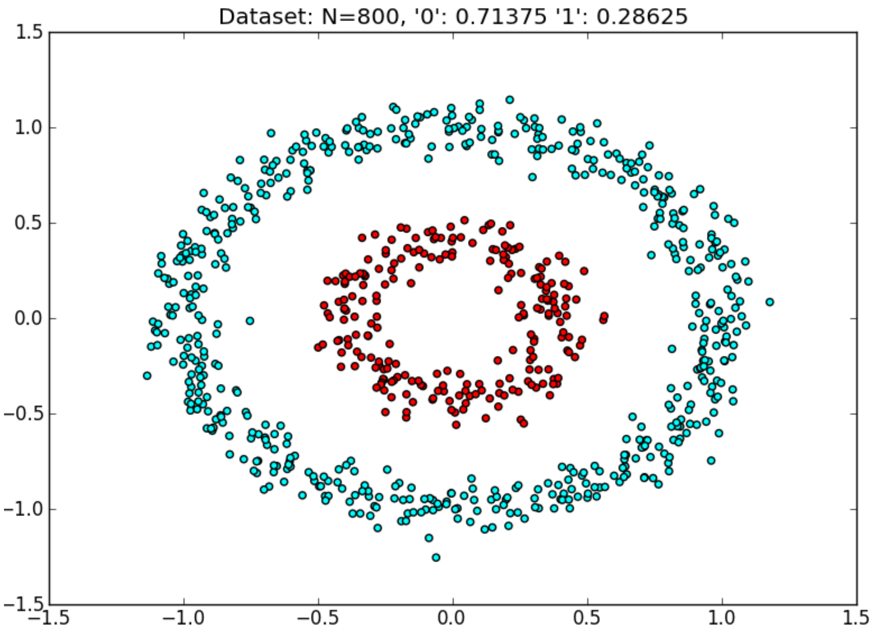
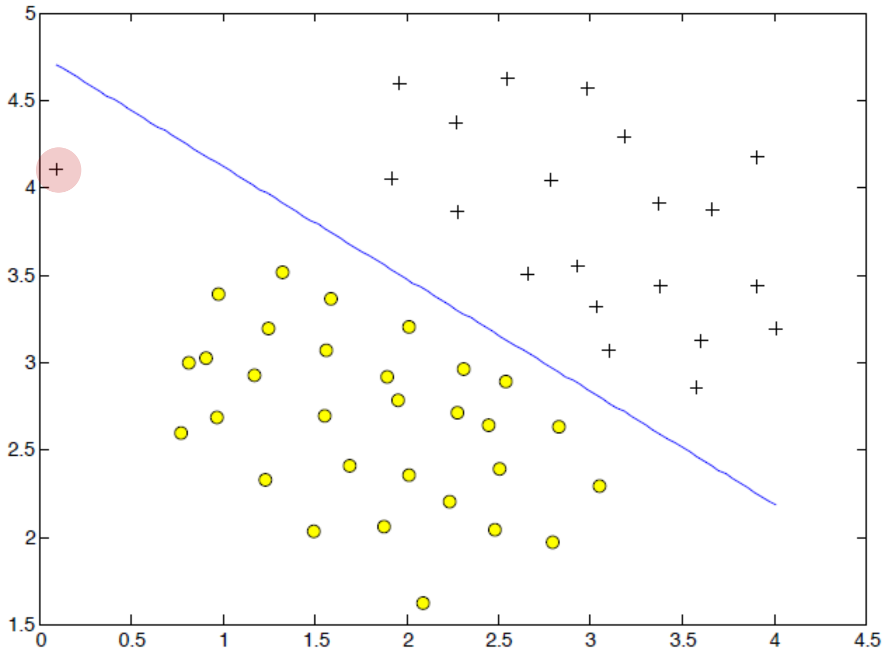


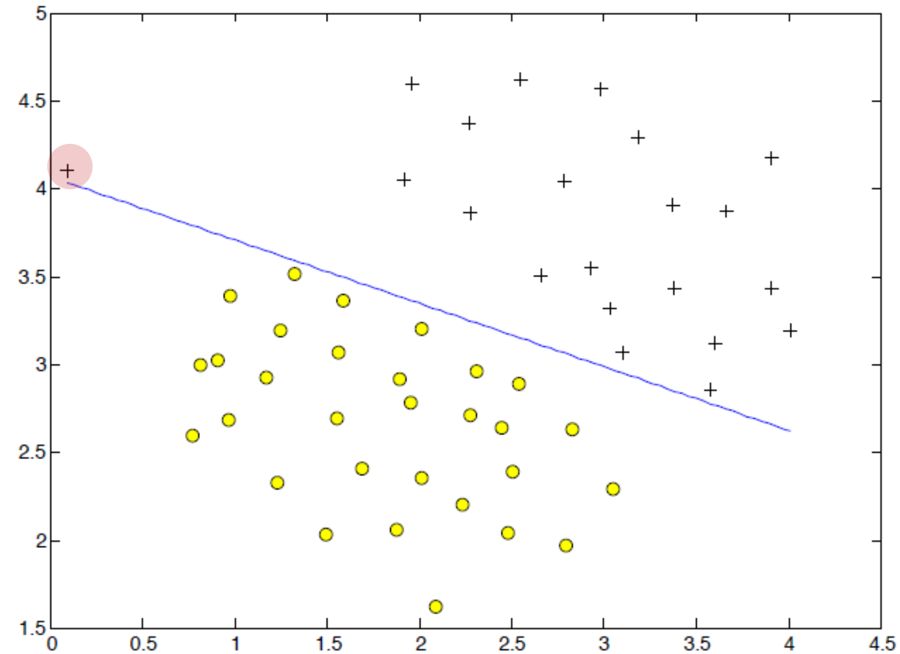



# Problemi koji NISU linearno separabilni

## Relaksacija definicije „razdvajanja“

# Soft margin SVM – motivacija




- Do sada smo razmatrali *hard margin* SVM – granica odluke je idealno razdvajala skup podataka na pozitivne i negativne primere
- Sada ćemo razmotriti *soft margin* SVM – dozvolićemo „greške“
- Ne želimo da *outlieri* diktiraju granicu odluke

# Soft margin SVM

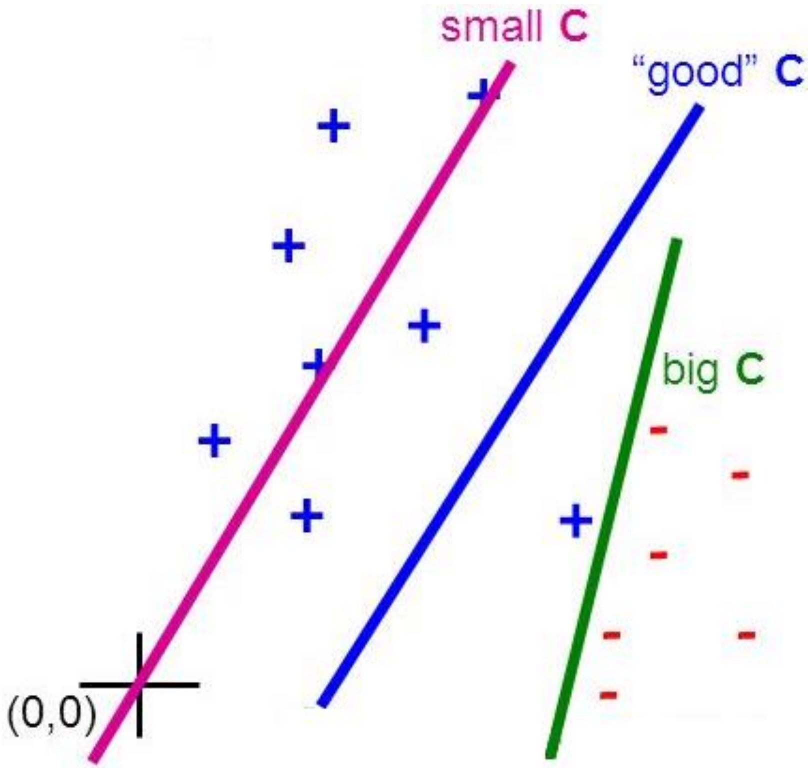
- Dozvolićemo greške
- Promena optimizacionog problema:

$$\min \frac{1}{2}\theta^T\theta + C \cdot (\text{br. grešaka})$$

- Dakle, minimizujemo $\theta^T\theta$ plus broj grešaka

- Kako da odaberemo $C$ (*slack penalty*)?
  - Veliko $C$ → jako kažnjavamo greške
  - Malo $C$ → dozvoljavamo mnogo grešaka

# Kako da odaberemo $C$ (*slack penalty*)?




$$\min \frac{1}{2}\theta^T\theta + C \cdot (\text{br. grešaka})$$

- Veliko $C$ → jako kažnjavamo greške
- Malo $C$ → nije nam uopšte važno da se prilagodimo podacima, već samo da $\theta$ bude 0

# Kako da odaberemo $C$ (*slack penalty*)?

- $C$ – kontroliše nagodbu između širine margine i koliko kažnjavamo greške

- Igra ulogu sličnu $\lambda$ u logističkoj regresiji $\left(C = \frac{1}{\lambda}\right)$
  - $C = \infty$ – naš jedini cilj je da želimo $\theta$ i $b$ koji razdvajaju podatke (overfitting)
  - $C = 0$ – nije nas briga za greške, želimo što manji $\theta$ pa će ispasti $\theta = 0$. Praktično, ignorišemo podatke (underfitting)

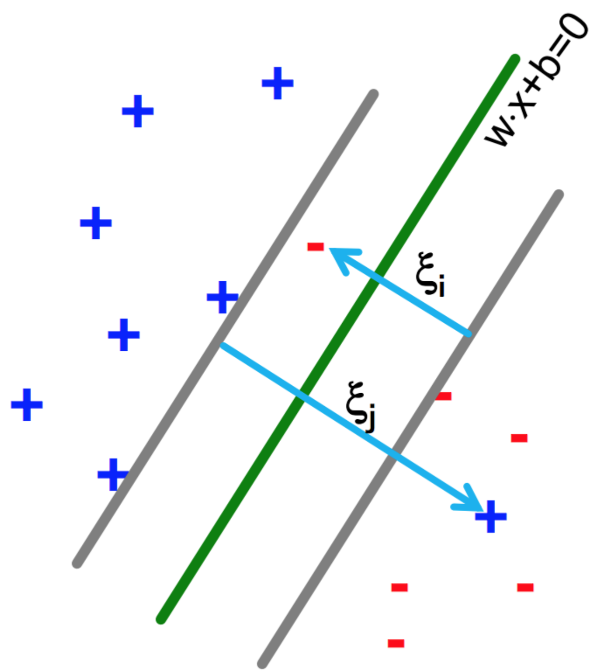
- Optimalno $C$: unarksna validacija ili korišćenje posebnog validacionog skupa

# Soft margin SVM

- Promena optimizacionog problema:

$$\min \frac{1}{2}\theta^T\theta + C \cdot (\text{br. grešaka})$$

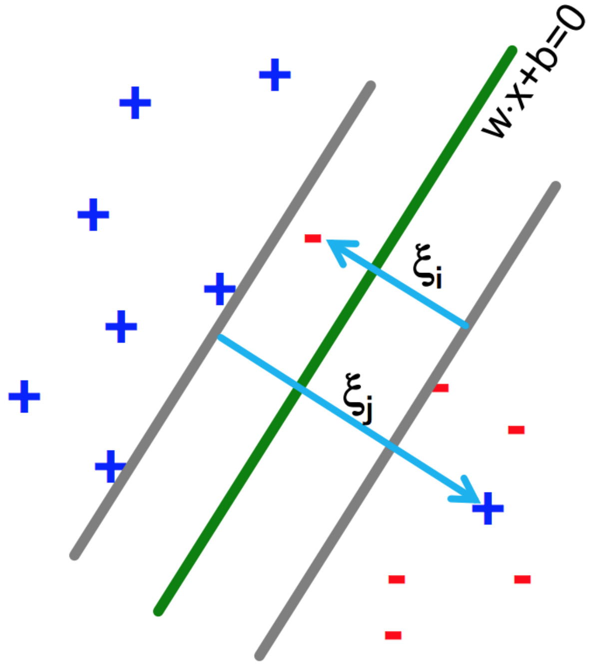
- Ali nisu sve greške jednako loše!




- Uvešćemo *slack variables* $\xi_i$
  - $\xi_i$ je „kazna“ za datu grešku
  - definisaćemo je kao rastojanje druge strane margine do date tačke

# Soft margin SVM

| <u>Osnovna formulacija</u> | <u>Linearno neseparabilan slučaj</u> |
|---|---|
| $\min \frac{1}{2}\theta^T\theta$ | $\min \frac{1}{2}\theta^T\theta + C\left(\sum_{n=1}^{N}\xi_i\right)$ |
| $y^{(i)}\left(\theta^T x^{(i)} + b\right) \geq 1 \text{ za } i = 1,2,\dots,N$ | $y^{(i)}\left(\theta^T x^{(i)} + b\right) \geq 1 - \xi_n \text{ za } i = 1,2,\dots,N$ |




- U slučaju da podaci nisu linearno separabilni, potrebno je dozvoliti greške, ali uz težnju da se njihov broj i intenzitet minimizuju

- Konstanta $C$ kontroliše koliko je margina „relaksirana“

# Hinge loss

$$\min \frac{1}{2}\overbrace{\theta^T\theta}^{\text{Margina}} + \underbrace{C}_{\text{Regularizacioni parametar}} \left( \overbrace{\sum_{n=1}^{N} \xi_i}^{\text{Empirical loss (koliko dobro fitujemo podatke)}} \right)$$

$$y^{(i)}\left(\theta^T x^{(i)} + b\right) \geq 1 - \xi_i \ \text{ za } i = 1,2, \dots, N$$

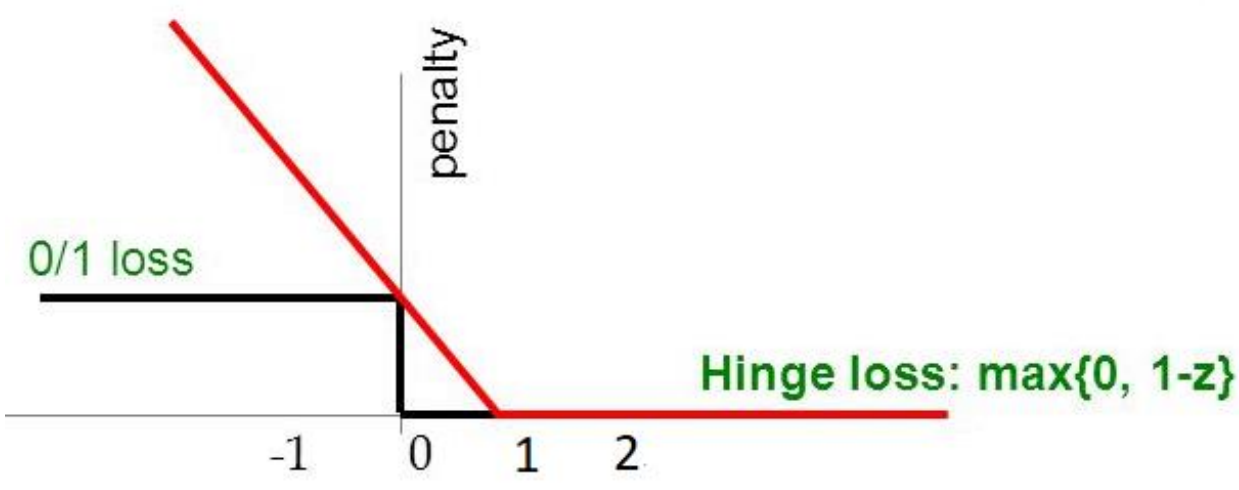
$$\xi_i \geq 0 \text{ za } \ i = 1,2, \dots, N$$

- Ovaj minimizacioni problem možemo preformulisati:

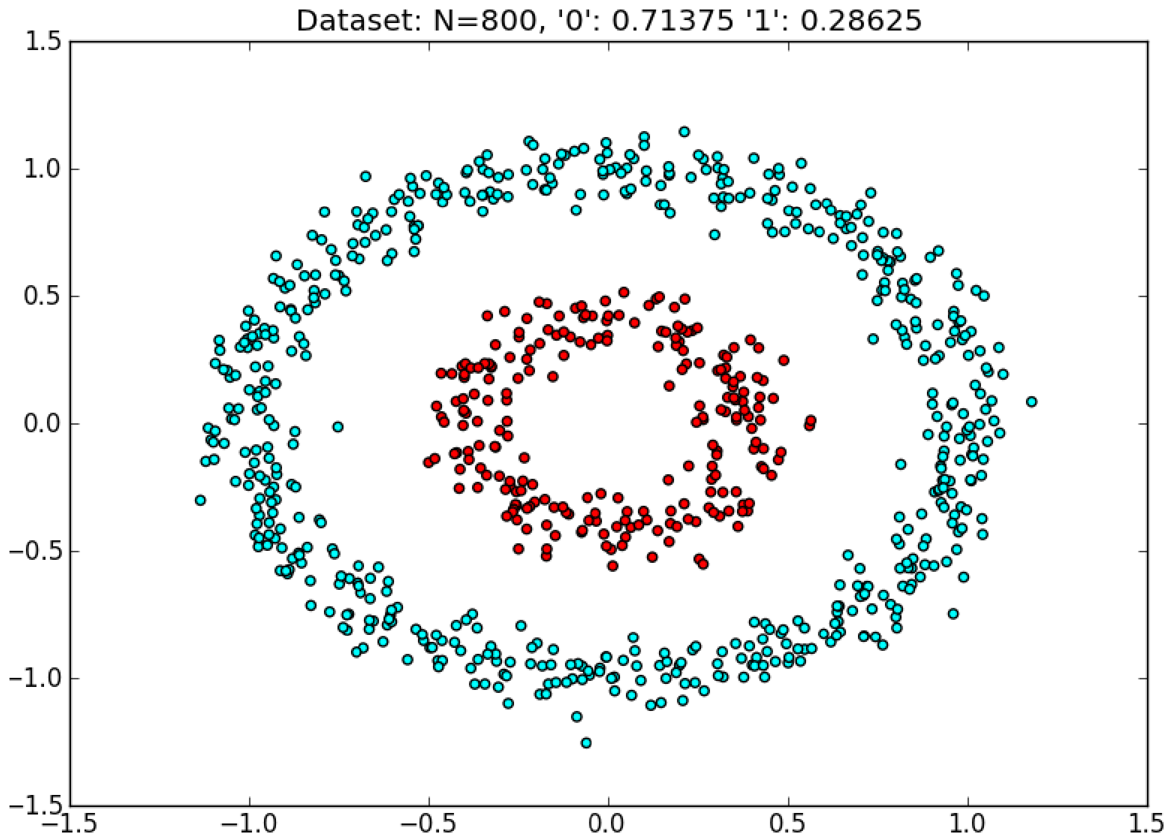
$$\left[\sum_{i=1}^{N} \max\left(0,1 - y^{(i)}\left(\theta^T x^{(i)} + b\right)\right)\right] + \lambda \frac{\theta^T\theta}{2}$$

# Hinge loss

- SVM koristi *hinge loss* (greška u obliku šarke)
- 0/1 loss: ako dobro klasifikujemo, „kazna“ je 0, a ako pogrešno klasifikujemo, „kazna“ je 1
- *Hinge loss*:
  - Ako dobro klasifikujemo tačku, i rastojanje tačke od granice odluke je barem 1, „kazna“ je 0
  - Ako je tačka klasifikovana korektno, ali je preblizu granice odluke) ili je pogrešno klasifikovana, „kazna“ je proporcionalna udaljenosti tačke od granice odluke



# Možemo li na ovo primeniti soft margin?




- Ne postoji linija koja na dobar način razdvaja crvenu i plavu klasu
- SVM ima ograničenje da granica odluke mora da bude linearna
  - U izvođenjima podrazumevamo da je granica odluke razdvajajuća **hiperravan**

# Šta ako problem nije linearno separabilan?

- Rešavamo ga kombinacijom dva načina:
  - Relaksiranjem naše definicije „razdvajanja“
  - ➢Nelinearnim transformacijama podataka (kernel trik)