# Hilbertrum med Reproducerande Kärnor

Oscar Granlund

13 april 2018

**Sammanfattning**

Testtesttesttesttesttest

# Kapitel 1

# Stödvektormaskiner (SVM)

## 1.1 Klassificering med hjälp av separerande hyperplan

INTRODUKTION OM VARFÖR KLASSIFICERING, EXEMPEL MED SPAM-FILTER

**Definition 1.1.1.** Ett *klassificeringsproblem* är ett problem var man utgående från en mängd observationspar (*träningsdata*) $(\mathbf{x}_i,\ y_i)$, $\mathbf{x}_i \in \mathbb{R}^p$, $y_i \in \{-1,\ 1\}$, $i = 1, \dots,\ N$, försöker hitta en regel $g : \mathbb{R}^p \longmapsto \{-1,\ 1\}$ sådan att $g(\mathbf{x}_i) = y_i$ för alla träningspar $(\mathbf{x}_i,\ y_i)$.

Inom statistiken och maskininlärningen finns många olika metoder för att försöka lösa klassificeringsproblem, till exempel med hjälp av regression eller någon sorts klusteralgoritm. I detta kapitel behandlas en metod där affina mängder med dimensionerna $p - 1$ används för att definiera en regel som klassificerar *observationerna* $\mathbf{x}_i$ i *klasserna* $y_i \in \{-1,\ 1\}$ genom separering.

**Definition 1.1.2.** Ett *hyperplan* i ett vektorrum med dimensionen $p$ är ett underrum med dimensionen $p-1$; figur 1.1 illustrerar ett separerande hyperplan för fallet $p = 2$. Klassificeringsregeln $g$ för separerande hyperplan blir $g(\mathbf{x}_i) = \text{sign}(\mathbf{x}_i^\intercal \boldsymbol{\beta} + \beta_0)$ där mängden $\{\mathbf{x} : \mathbf{x}^\intercal \boldsymbol{\beta} + \beta_0 = 0\}$, med $\mathbf{x},\ \boldsymbol{\beta} \in \mathbb{R}^p$ och $\|\boldsymbol{\beta}\| = 1$, definierar ett hyperplan, eller en *affin* mängd, parametriserat av $\boldsymbol{\beta}$ och $\beta_0$.

**Sats 1.1.1.** Ett hyperplan definierat som den affina mängden $L = \{\mathbf{x} : f(\mathbf{x}) = \mathbf{x}^\intercal \boldsymbol{\beta} + \beta_0 = 0\}$ har följande egenskaper [1]:

1. Den normaliserade normalvektorn $\widehat{\boldsymbol{\beta}}$ kan skrivas på formen

$$\widehat{\boldsymbol{\beta}} = \frac{\boldsymbol{\beta}}{\|\boldsymbol{\beta}\|}.$$

2. $\mathbf{x}_0^\intercal \boldsymbol{\beta} = -\beta_0$ för alla $\mathbf{x}_0$ i $L$.

3. Det signerade avståndet från en punkt $\mathbf{x}$ till hyperplanet $L$ ges av

$$\begin{aligned}(\mathbf{x} - \mathbf{x}_0)^\intercal \widehat{\boldsymbol{\beta}} &= \frac{1}{\|\boldsymbol{\beta}\|}(\mathbf{x}^\intercal \boldsymbol{\beta} + \beta_0) \\ &= \frac{1}{\|f'(\mathbf{x})\|} f(\mathbf{x}).\end{aligned}$$

*Bevis.*

1. Låt $\mathbf{x}_1$ och $\mathbf{x}_2$ vara två punkter i $L$. Då gäller att $f(\mathbf{x}_1) = f(\mathbf{x}_2) = 0$ och

$$\begin{aligned}0 &= f(\mathbf{x}_1) - f(\mathbf{x}_2) \\ &= \mathbf{x}_1^\intercal \boldsymbol{\beta} + \beta_0 - \mathbf{x}_2^\intercal \boldsymbol{\beta} - \beta_0 \\ &= (\mathbf{x}_1 - \mathbf{x}_2)^\intercal \boldsymbol{\beta}\end{aligned}$$

alltså uppfyller $\boldsymbol{\beta}$ kravet för normalvektorer och $\widehat{\boldsymbol{\beta}} := \frac{\boldsymbol{\beta}}{\|\boldsymbol{\beta}\|}$ är den normaliserade normalvektorn till hyperplanet $L$. ∎

2. Låt $\mathbf{x}_0$ vara en punkt i $L$. Då gäller att $f(\mathbf{x}_0) = \mathbf{x}_0^\intercal \boldsymbol{\beta} + \beta_0 = 0$ alltså är $\mathbf{x}^\intercal \boldsymbol{\beta} = -\beta_0$. ∎

3. Låt $\mathbf{x}_0$ vara en punkt i hyperplanet $L$. Då är avståndet från hyperplanet till punkten $\mathbf{x}$ lika med längden av projektionen av vektorn $(\mathbf{x} - \mathbf{x}_0)$ på hyperplanets normal, $\beta$. Vi får alltså att

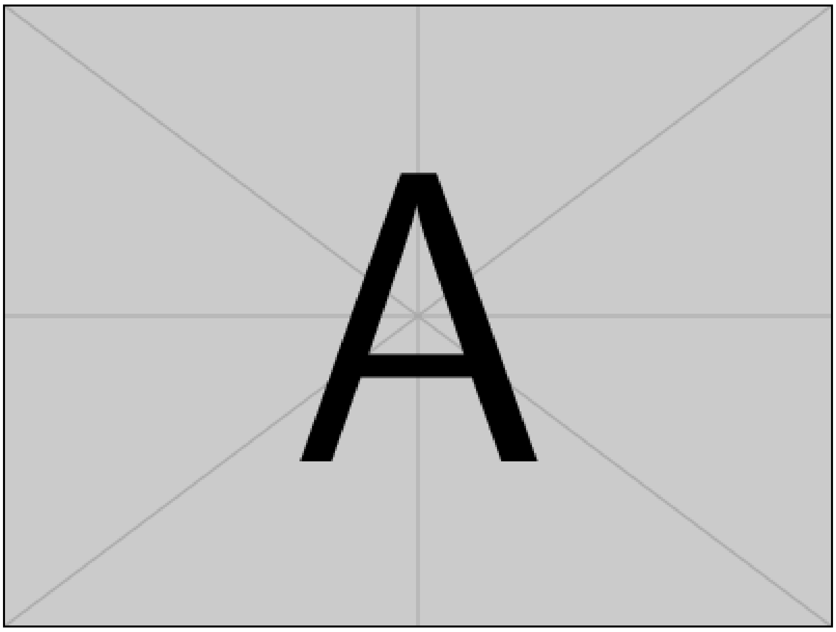
$$\begin{aligned}\mathrm{d}^{\pm}(\mathbf{x}, L) &= \mathrm{comp}_{\boldsymbol{\beta}}(\mathbf{x} - \mathbf{x}_0) = \underline{\underline{(\mathbf{x} - \mathbf{x}_0)^\intercal \widehat{\boldsymbol{\beta}}}} \\ &= \frac{1}{\|\boldsymbol{\beta}\|}(\mathbf{x}^\intercal \boldsymbol{\beta} - \mathbf{x}_0^\intercal \boldsymbol{\beta}) = \underline{\underline{\frac{1}{\|\boldsymbol{\beta}\|}(\mathbf{x}^\intercal \boldsymbol{\beta} + \boldsymbol{\beta}_0)}}\end{aligned}$$

och om man noterar att $f(\mathbf{x}) = \mathbf{x}^\intercal \boldsymbol{\beta} + \beta_0$ och $f'(\mathbf{x}) = \boldsymbol{\beta}$ så fås även att

$$\frac{1}{\|\boldsymbol{\beta}\|}(\mathbf{x}^\intercal \boldsymbol{\beta} + \boldsymbol{\beta}_0) = \frac{1}{\|f'(\mathbf{x})\|} f(\mathbf{x}).$$

∎

*Observation.* Definitionen för hyperplanet $L = \{\mathbf{x} : f(\mathbf{x}) = \mathbf{x}^\intercal \boldsymbol{\beta} + \beta_0 = 0\}$ är inte entydig.




Figur 1.1: 20 datapunkter med ett separerande hyperplan (linje) där klassen $y = 1$ har färgats blå och klassen $y = -1$ har färgats orange.

*Orsak.* Betrakta hyperplanen $L_1 = \{\mathbf{x} : f(\mathbf{x}) = \mathbf{x}^\intercal \boldsymbol{\beta} + \beta_0 = 0\}$ och $L_2 = \{\mathbf{x} : g(\mathbf{x}) = \mathbf{x}^\intercal(-1 \cdot \boldsymbol{\beta}) + (-1 \cdot \beta_0)\}$. Eftersom att $g(\mathbf{x}) = -f(\mathbf{x})$ så gäller att om $\mathbf{x}$ tillhör $L_1$ så tillhör $\mathbf{x}$ även $L_2$. Betrakta vidare $L_3 = \{\mathbf{x} : h(\mathbf{x}) = \frac{\mathbf{x}^\intercal \boldsymbol{\beta}}{\|\boldsymbol{\beta}\|} + \frac{\beta_0}{\|\boldsymbol{\beta}\|}\} = 0$. Om $\mathbf{x}$ då tillhör $L_1$ så tillhör $\mathbf{x}$ även $L_3$ eftersom att $h(\mathbf{x}) = \frac{f(\mathbf{x})}{\|\boldsymbol{\beta}\|} = 0$. Notera även att $\|\boldsymbol{\beta}\|$ kunde ha varit vilket reellt tal $\alpha$ som helst.

*Observation.* För att få entydiga hyperplan för klassificering kan man lägga till villkor. Om man kräver att $\|\boldsymbol{\beta}\| = 1$ och $y_i(\mathbf{x}_i^\intercal \boldsymbol{\beta} + \beta_0) \geq 0$ för alla $i = 1, \ldots,\ N$, där $y_i$ är klasserna i klassificeringsproblemet, så får man en entydig definition av hyperplanet där vektorn $\boldsymbol{\beta}$ "pekar mot" klassen där $y_i = 1$ och $\beta_0$ anger det signerade avståndet (med avseende på vart $\boldsymbol{\beta}$ pekar) från origo till hyperplanet.

*Orsak.* De extra villkoren gör att man inte längre kan göra manipulationerna som påvisade icke-entydigheten. Om man sätter $\mathbf{x} = \bar{0}$ så får man med hjälp av sats 1.1.1 att avståndet från origo till planet är lika med $\frac{1}{\|\boldsymbol{\beta}\|}(\mathbf{x}^\intercal \boldsymbol{\beta} + \beta_0) = \beta_0$.

**Definition 1.1.3.** Ett klassificeringsproblem kallas *separabelt* om det existerar ett hyperplan $L = \{\mathbf{x} : f(\mathbf{x}) = \mathbf{x}^\intercal \boldsymbol{\beta} + \beta_0 = 0\}$ som separerar mängderna.

**Sats 1.1.2.** För ett hyperplan $L = \{\mathbf{x} : f(\mathbf{x}) = \mathbf{x}^\intercal \boldsymbol{\beta} + \beta_0 = 0,\ y_i f(\mathbf{x}_i) \geq$

$0,\ i = 1, \ldots,\ N,\ \|\boldsymbol{\beta}\| = 1\}$ som separerar två klasser gäller att

$$y_i(\mathbf{x}_i^\intercal \boldsymbol{\beta} + \beta_0) > 0 \tag{1.1}$$

för alla $i = 1, \ldots,\ N$.

*Bevis.* Ifall ett klassificeringsproblem är separabelt så ligger alla observationer $y_i$ på rätt sida av hyperplanet definierat genom $\mathbf{x}^\intercal \boldsymbol{\beta} + \beta_0$; eller så ligger alla observationer på fel sida av hyperplanet. Vilket betyder att ifall $y_i = 1$ så är $\mathbf{x}_i^\intercal \boldsymbol{\beta} + \beta_0 > 0$ och om $y_i = -1$ så är $\mathbf{x}_i^\intercal \boldsymbol{\beta} + \beta_0 < 0$. Detta betyder att $y_i(\mathbf{x}_i^\intercal \boldsymbol{\beta} + \beta_0) > 0$. Ifall $\mathbf{x}_i^\intercal \boldsymbol{\beta} + \beta_0 = 0$ är problemet inte separabelt. ■

**Exempel 1.1.1.** Låt träningsdataparen vara $([2,\ 2]^\intercal,\ 1),\ ([1,\ 2]^\intercal, -1)$. Då är

$$L_1 = \{\mathbf{x} \in \mathbb{R}^2 : \mathbf{x}^\intercal \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 1.5 = 0\}$$

och

$$L_2 = \{\mathbf{x} \in \mathbb{R}^2 : \mathbf{x}^\intercal \begin{bmatrix} \sqrt{2} \\ \sqrt{2} \end{bmatrix} + 1.5 = 0\}$$

två separerande hyperplan (linjer i detta fall).

*Bevis.* För $L_1$: ■

# Litteraturförteckning

[1] Hastie Trevor, Robert Tibshirani, and Jerome Friedman. *The elements of statistical learning : data mining, inference, and prediction.* Springer series in statistics. Springer New York Inc., 2001.