Hilbertrum med Reproducerande Kärnor

Oscar Granlund

1 augusti 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*) $(\mathbf{x}_i, y_i)$, $\mathbf{x}_i \in \mathbb{R}^p$, $y_i \in \{-1, 1\}$, $i = 1, \ldots, N$, 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 $\mathbf{x},\ \boldsymbol{\beta} \in \mathbb{R}^p$, definierar ett hyperplan alternativt en *affin* mängd, parametriserat av $\boldsymbol{\beta}$ och β_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 $\widehat{\boldsymbol{\beta}}_{\text{n}}$ kan skrivas på formen

$$\widehat{\boldsymbol{\beta}}_{\text{n}} = \frac{\boldsymbol{\beta}}{\|\boldsymbol{\beta}\|}.$$

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

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

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 $\widehat{\boldsymbol{\beta}}_{\mathrm{n}} := \frac{\boldsymbol{\beta}}{\|\boldsymbol{\beta}\|}$ 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, β. Vi får alltså att

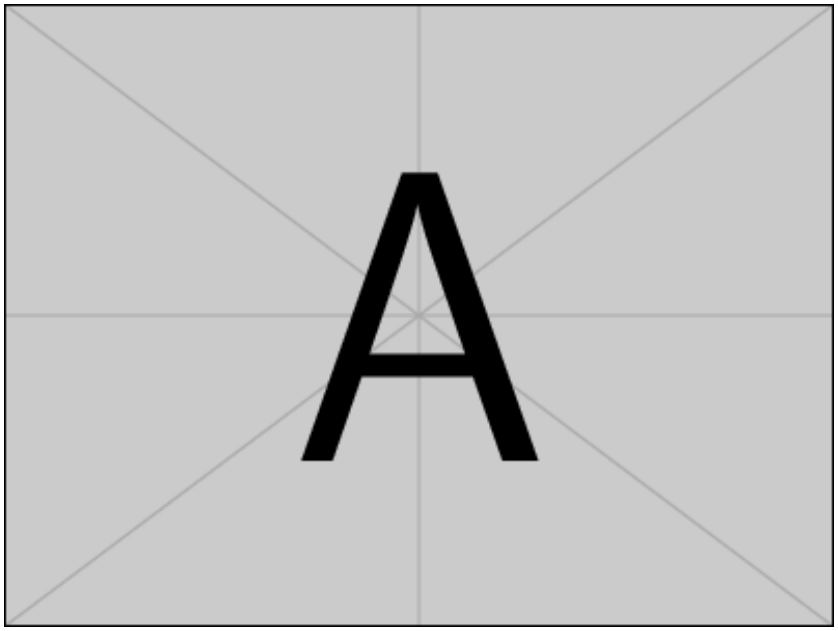
$$\begin{aligned}\mathrm{d}^{\pm}(\mathbf{x}, L) = \mathrm{comp}_{\boldsymbol{\beta}}(\mathbf{x} - \mathbf{x}_0) &= \underline{\underline{(\mathbf{x} - \mathbf{x}_0)^\intercal \widehat{\boldsymbol{\beta}}}} \\ = \frac{1}{\|\boldsymbol{\beta}\|}(\mathbf{x}^\intercal \boldsymbol{\beta} - \mathbf{x}_0^\intercal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

   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 α 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 β_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 *linjärt 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 linjärt 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 linjärt 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

$$L_2 = \{\mathbf{x} \in \mathbb{R}^2 : \mathbf{x}^\intercal \begin{bmatrix} \sqrt{2} \\ \sqrt{2} \end{bmatrix} - 3.5\sqrt{2} = 0\}$$

två separerande hyperplan (linjer i detta fall).

Bevis. För L_1:

$$y_1(\mathbf{x}_1^\intercal \begin{bmatrix} 1 \\ 0 \end{bmatrix} - 1.5) = [2,\ 2]^\intercal \begin{bmatrix} 1 \\ 0 \end{bmatrix} - 1.5 = 0.5 > 0$$

och

$$y_2(\mathbf{x}_2^\intercal \begin{bmatrix} 1 \\ 0 \end{bmatrix} - 1.5) = -1([1,\ 2]^\intercal \begin{bmatrix} 1 \\ 0 \end{bmatrix} - 1.5) = (-1)(-0.5) = 0.5 > 0.$$

Och för L_2:

$$y_1(\mathbf{x}_1^\intercal \begin{bmatrix} \sqrt{2} \\ \sqrt{2} \end{bmatrix} - 3.5\sqrt{2}) = [2,\ 2]^\intercal \begin{bmatrix} \sqrt{2} \\ \sqrt{2} \end{bmatrix} - 3.5\sqrt{2} = 0.5\sqrt{2} > 0$$

och

$$y_2(\mathbf{x}_2^\intercal \begin{bmatrix} \sqrt{2} \\ \sqrt{2} \end{bmatrix} - 3.5\sqrt{2}) = -1([1,\ 2]^\intercal \begin{bmatrix} \sqrt{2} \\ \sqrt{2} \end{bmatrix} - 3.5\sqrt{2}) = (-1)(-0.5\sqrt{2}) = 0.5\sqrt{2} > 0$$

■

Observation. Hyperplan kan konstrueras enkelt genom att man i $\mathbb{R}^n$ väljer n stycken punkter $\mathbf{x}_i$ som man vill att planet ska gå igenom och sedan löser ekvationssystemet $X\boldsymbol{\beta} = -\boldsymbol{\beta}_0$ där X är en matris där raderna består av punkterna $\mathbf{x}_i$, $i = 1, \ldots, n$ och $\boldsymbol{\beta}_0$ är en vektor av värdena β_0.

Som syns i exempel 1.1.1 så kan det finnas många separerande hyperplan ifall ett klassificeringsproblem är linjärt separabelt och frågan är ju då vilket av alla separerande hyperplan man borde välja.

1.2 Optimala separerande hyperplan

Inom statistiken finns många olika sätt att anpassa en modell till data och dessa sätt kan ofta visas vara ekvivalenta med något optimeringsproblem, till exempel maximum likelihood-metoden (ML-metoden) för linjär regression som kan visas vara ekvivalent med minstakvadratmetoden. Dessa optimeringsproblem kan oftast ändras genom att man lägger till eller tar bort termer i objektivfunktionen eller ändrar på kraven.

För separerande hyperplan kommer vi att behandla ett optimeringsproblem som är utformat så att det kortaste avståndet från hyperplanet till de närmaste träningspunkterna från vardera klass maximeras [**?**]. Med andra ord fås följande optimeringsproblem

$$\begin{aligned} &\max_{\widehat{\boldsymbol{\beta}},\beta_0,\|\widehat{\boldsymbol{\beta}}\|=1} && C \\ &\text{så att} && y_i(\mathbf{x}_i^\intercal \widehat{\boldsymbol{\beta}} + \beta_0) \geq C, \quad i = 1, \ldots, N \end{aligned} \tag{1.2}$$

där C kallas marginalen och betecknar avståndet från hyperplanet till de närmaste punkterna.

Observation. Ifall alla punkter är rätt klassificerade så anger $y_i(\mathbf{x}_i^\intercal \widehat{\boldsymbol{\beta}} + \beta_0)$ det absoluta avståndet mellan hyperplanet och punkten $\mathbf{x}_i$.

Förhoppningen är här att man genom att välja det separerande hyperplan som befinner sig så långt som möjligt från båda klasserna hittar ett hyperplan som även generaliserar till ny data så bra som möjligt. Dessutom kommer vi också se att detta är ett sätt att unikt välja ett separerande hyperplan.

För att se att optimeringsproblemet (1.2) är *konvext*, det vill säga har en unik lösning, måst vi skriva om det något. Vi börjar med att gör oss av med kravet $\|\widehat{\boldsymbol{\beta}}\| = 1$ genom att byta ut kraven

$$y_i(\mathbf{x}_i^\intercal \widehat{\boldsymbol{\beta}} + \beta_0) \geq C, \quad i = 1, \ldots, N$$

mot kraven

$$y_i(\mathbf{x}_i^\intercal \frac{\boldsymbol{\beta}}{\|\boldsymbol{\beta}\|} + \frac{\beta_1}{\|\boldsymbol{\beta}\|}) = \frac{1}{\|\boldsymbol{\beta}\|} y_i(\mathbf{x}_i^\intercal \boldsymbol{\beta} + \beta_1) \geq C, \quad i = 1, \ldots, N$$

eller ekvivalent

$$y_i(\mathbf{x}_i^\intercal \boldsymbol{\beta} + \beta_1) \geq C\|\boldsymbol{\beta}\|, \quad i = 1, \ldots, N.$$

där man alltså valt en av de andra representationerna för samma hyperplan genom att skala om $\widehat{\boldsymbol{\beta}}$ och β_0. Vidare kan C elimineras genom att man väljer $C = \frac{1}{\|\boldsymbol{\beta}\|}$ och får

$$y_i(\mathbf{x}_i^\intercal \boldsymbol{\beta} + \beta_1) \geq 1, \quad i = 1, \ldots, N$$

och eftersom att $C = \frac{1}{\|\boldsymbol{\beta}\|}$ ökar när $\|\boldsymbol{\beta}\|$ minskar är maximering av C ekvivalent med minimering av $\|\boldsymbol{\beta}\|$ så vi får optimeringsproblemet

$$\begin{aligned} \min_{\boldsymbol{\beta},\beta_1} \quad & \|\boldsymbol{\beta}\| \\ \text{så att} \quad & y_i(\mathbf{x}_i^\intercal \boldsymbol{\beta} + \beta_1) \geq 1, \quad i = 1, \ldots, N. \end{aligned}$$

Därefter görs ännu en kvadratisk transformering av *kostfunktionen* $\|\boldsymbol{\beta}\|$ alltså man noterar att $\operatorname{argmin}_{\boldsymbol{\beta},\beta_1} \|\boldsymbol{\beta}\| = \operatorname{argmin}_{\boldsymbol{\beta},\beta_1} \frac{1}{2}\|\boldsymbol{\beta}\|^2$. Med andra ord har vi ett optimeringsproblem av en kvadratisk funktion med linjära krav alltså ett konvext optimeringsproblem där lösningar existerar.

Observation. För två vektorer $\mathbf{a}$ och $\mathbf{b}$ i $\mathbb{R}^n$ kan produkten $\mathbf{a}^\intercal\mathbf{b}$ uttryckas som den normala inre produkten $\langle \mathbf{a}, \mathbf{b} \rangle$ i $\mathbb{R}^n$. Detta kommer att komma till nytta i kapitel 2 där konvexiteten för en utvidgning av det linjära problemet utforskas.

Ovanstående resonemang ger alltså ett bevis för sats 1.2.1.

Sats 1.2.1. Låt $\widehat{\boldsymbol{\beta}},\ \boldsymbol{\beta} \in \mathbb{R}^p$ och $\beta_0,\ \beta_1 \in \mathbb{R}$. Då är optimeringsproblemet

$$\begin{aligned} \max_{\widehat{\boldsymbol{\beta}},\beta_0,\|\widehat{\boldsymbol{\beta}}\|=1} \quad & C \\ \text{så att} \quad & y_i(\mathbf{x}_i^\intercal \widehat{\boldsymbol{\beta}} + \beta_0) \geq C, \quad i = 1, \ldots, N \end{aligned}$$

konvext och ekvivalent med optimeringsproblemet

$$\begin{aligned} \min_{\boldsymbol{\beta},\beta_1} \quad & \frac{1}{2}\|\boldsymbol{\beta}\|^2 \\ \text{så att} \quad & y_i(\mathbf{x}_i^\intercal \boldsymbol{\beta} + \beta_1) \geq 1, \quad i = 1, \ldots, N \end{aligned}$$

ifall träningsdatat $(\mathbf{x}_i, y_i)$ är sådant att klassificeringsproblemet är linjärt separabelt.

Vidare ger observationen ovan ett till ekvivalent optimeringsproblem:

Korollarium 1.2.2. Optimeringsproblemen i sats 1.2.1 är ekvivalenta med optimeringsproblemet

$$\begin{aligned} \min_{\boldsymbol{\beta},\beta_1} \quad & \frac{1}{2}\langle \boldsymbol{\beta}, \boldsymbol{\beta} \rangle \\ \text{så att} \quad & y_i(\langle \mathbf{x}_i, \boldsymbol{\beta} \rangle + \beta_1) \geq 1, \quad i = 1, \ldots, N \end{aligned}$$

för samma krav.

Bevis. Normen $\|\boldsymbol{\beta}\| = (\boldsymbol{\beta}^\intercal \boldsymbol{\beta})^{\frac{1}{2}}$ kan uttryckas som $\langle \boldsymbol{\beta}, \boldsymbol{\beta} \rangle^{\frac{1}{2}}$ alltså är $\frac{1}{2}\|\boldsymbol{\beta}\|^2 = \frac{1}{2}\langle \boldsymbol{\beta}, \boldsymbol{\beta} \rangle$. Resten följer från observationen. ■

Framställningen i korollarium 1.2.2 används i många källor, bland annat i den ursprungliga framställningen för stödvektormaskinen och är en av de mer generella framställningarna för optimeringsproblemet som stödvektormaskinen bygger på.

För att hitta alla extrempunkter till ett optimeringsproblem, det vill säga lösa ett konvext optimeringsproblem, används Lagrangemultiplikatorer. Den primala Lagrangefunktionen L_P för optimeringsproblemet

$$\begin{aligned} \min_{\boldsymbol{\beta},\beta_1} \quad & \frac{1}{2}\|\boldsymbol{\beta}\|^2 \\ \text{så att} \quad & y_i(\mathbf{x}_i^\intercal \boldsymbol{\beta} + \beta_1) \geq 1, \quad i = 1, \ldots, N \end{aligned}$$

ges då av

$$L_P = \frac{1}{2}\|\boldsymbol{\beta}\|^2 - \sum_{i=1}^{N} \lambda_i [y_i(\mathbf{x}_i^\intercal \boldsymbol{\beta} + \beta_1) - 1] \tag{1.3}$$

som ska minimeras med avseende på $\boldsymbol{\beta}$ och β_1.

För att minimera L_P sätter vi derivatorna med avseende på elementen $[\boldsymbol{\beta}]_j$ av $\boldsymbol{\beta}$ och β_1 till 0, det vill säga vi får följande relationer:

$$\begin{aligned} D_{[\boldsymbol{\beta}]_j}(L_P) &= D_{[\boldsymbol{\beta}]_j}\left(\frac{1}{2}\boldsymbol{\beta}^\intercal \boldsymbol{\beta}\right) && - D_{[\boldsymbol{\beta}]_j}\left(\sum_{i=1}^{N} \left(\lambda_i y_i \left(\mathbf{x}_i^\intercal \boldsymbol{\beta}\right) + \lambda_i y_i \beta_1 - \lambda_i\right)\right) \\ &= D_{[\boldsymbol{\beta}]_j}\left(\frac{1}{2}\sum_{k=1}^{p} [\boldsymbol{\beta}]_k^2\right) && - \sum_{i=1}^{N} D_{[\boldsymbol{\beta}]_j}\left(\lambda_i y_i \left(\sum_{k=1}^{p} [\mathbf{x}_i]_k [\boldsymbol{\beta}]_k\right) + \lambda_i y_i \beta_1 - \lambda_i\right) \\ &= [\boldsymbol{\beta}]_j && - \sum_{i=1}^{N} D_{[\boldsymbol{\beta}]_j}\left(\sum_{k=1}^{p} \lambda_i y_i [\mathbf{x}_i]_k [\boldsymbol{\beta}]_k\right) + 0 \\ &= [\boldsymbol{\beta}]_j && - \sum_{i=1}^{N} \lambda_i y_i [\mathbf{x}_i]_j \end{aligned} \tag{1.4}$$

där $j = 1, \ldots p$ och

$$D_{\beta_1}(L_P) = D_{\beta_1}\left(-\sum_{i=1}^{N} \lambda_i y_i \beta_1\right) = -\sum_{i=1}^{N} \lambda_i y_i.$$

Vidare kan relationen 1.4 skrivas om som derivatan med avseende på hela $\boldsymbol{\beta}$ eftersom att $[D_{\boldsymbol{\beta}}(L_p)]_j = D_{[\boldsymbol{\beta}]_j}(L_p)$. Efter att man tar i beaktande kraven att $D_{\boldsymbol{\beta}}(L_p) = \mathbf{0}$ och $D_{\beta_1}(L_p) = 0$ fås följande:

$$\boldsymbol{\beta} = \sum_{i=1}^{N} \lambda_i y_i \mathbf{x}_i$$
$$0 = \sum_{i=1}^{N} \lambda_i y_i.$$

1.2.1 Det duala problemet

Introduktion genom optimering (behövs kanske ett kapitel med optimeringsteori? Iaf. Lagrange-metoden)

1.3 Det oseparabla fallet

Optimera med slack-variabler

1.4 En enkel utvidgning med olinjära faktorer?

Visa att problemet lösbart

Kapitel 2

Hilbertrumteori, reproducerande kärnor

Varför utvidga fakoterna?

2.1 Grundläggande teori

Bevis av Mercers villkor för positivsemidefinita ekvationer/operatorer.

2.2 SVM som exempel

Något exempel. Introducera SVM för regression?

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.