KOÇ ÜNİVERSİTESİ

Çekirdeklerle öğrenme

Doç. Dr. Mehmet Gönen
mehmetgonen@ku.edu.tr

Endüstri Mühendisliği Bölümü, Mühendislik Fakültesi
Hesaplamalı Biyoloji Bölümü, Tıp Fakültesi

24 Haziran 2019

Çekirdek nedir?

- Çekirdek iki nesne arasındaki benzerlik miktarını hesaplayan bir fonksiyondur.
  - protein ikilileri
  - resim ikilileri
  - döküman ikilileri

- $\boldsymbol{x}_i \in \mathcal{X}$ ve $\boldsymbol{x}_j \in \mathcal{X}$
  - $\mathcal{X} =$ bütün olası proteinler
  - $\mathcal{X} =$ bütün olası resimler
  - $\mathcal{X} =$ bütün olası dökümanlar
- $k \colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$

Çekirdek nedir?

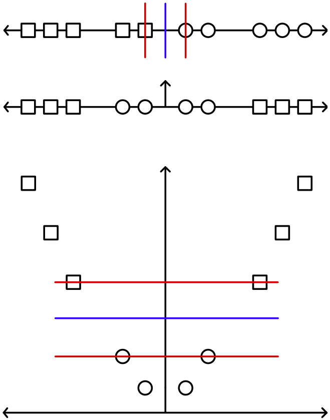
- Veri noktalarımızı genelde $N \times D$ büyüklüğünde bir matris olarak saklarız.

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1D} \\ x_{21} & x_{22} & \cdots & x_{2D} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{ND} \end{bmatrix}$$

- Vektör şeklinde gösterebildiğimiz girdiler için sıklıkla kullanılan çekirdekler:
  - $k_D(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{x}_i^\top \boldsymbol{x}_j$
  - $k_P(\boldsymbol{x}_i, \boldsymbol{x}_j) = (\boldsymbol{x}_i^\top \boldsymbol{x}_j + 1)^q$
  - $k_G(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp(-\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2 / s^2) = \exp(-d_{ij}^2 / s^2)$

Çekirdeklere neden ihtiyaç duyuyoruz?

- Doğrusal olmayan öznitelikler çıkarmakta kullanabiliriz.



Çekirdeklere neden ihtiyaç duyuyoruz?

- Her geçerli çekirdek fonksiyonu yeni bir öznitelik uzayına karşılık gelmektedir.
  $\Phi\colon \mathbb{R}^D \to \mathbb{R}^S$ (genelde $S \gg D$)
- Örneğin $D = 1$ ve $S = 3$
  $\Phi(x_i) = \begin{bmatrix} x_i^2 & \sqrt{2}x_i & 1 \end{bmatrix}^\top$
  $\Phi(x_i)^\top \Phi(x_j) = \begin{bmatrix} x_i^2 & \sqrt{2}x_i & 1 \end{bmatrix} \begin{bmatrix} x_j^2 & \sqrt{2}x_j & 1 \end{bmatrix}^\top = x_i^2 x_j^2 + 2x_i x_j + 1$
  $k_P(x_i, x_j) = (x_i x_j + 1)^2 = x_i^2 x_j^2 + 2x_i x_j + 1$
- 100 piksel $\times$ 100 piksel resimlerde ($D = 10^4$) aynısını yapabilir miyiz?
- Doğrusal olmayan öznitelikleri bu şekilde çıkarmak hesaplama maliyeti açısından uygulanabilir değil.

Çekirdeklere neden ihtiyaç duyuyoruz?

- $N \times D$ matris yerine $N \times N$ matris kullanarak hesaplama ve saklama maliyetini düşürebiliriz.

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1D} \\ x_{21} & x_{22} & \dots & x_{2D} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \dots & x_{ND} \end{bmatrix}$$

$$\mathbf{K} = \begin{bmatrix} k(\boldsymbol{x}_1, \boldsymbol{x}_1) & k(\boldsymbol{x}_1, \boldsymbol{x}_2) & \dots & k(\boldsymbol{x}_1, \boldsymbol{x}_N) \\ k(\boldsymbol{x}_2, \boldsymbol{x}_1) & k(\boldsymbol{x}_2, \boldsymbol{x}_2) & \dots & k(\boldsymbol{x}_2, \boldsymbol{x}_N) \\ \vdots & \vdots & \ddots & \vdots \\ k(\boldsymbol{x}_N, \boldsymbol{x}_1) & k(\boldsymbol{x}_N, \boldsymbol{x}_2) & \dots & k(\boldsymbol{x}_N, \boldsymbol{x}_N) \end{bmatrix}$$

Çekirdeklere neden ihtiyaç duyuyoruz?

- Belli bir yapıya sahip nesneler arasındaki benzerlik hesaplamakta kullanabiliriz.
- Örnek: İki protein arasındaki benzerlik

```
MVLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEK
FDRVKHLKTEAEMKASEDLKKHGVTVLTALGAILKKKGHHEAE
LKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRHPGNFGADAQ
GAMNKALELFRKDIAAKYKELGYQG
```

```
MNIFEMLRIDEGLRLKIYKDTEGYYTIGIGHLLTKSPSLNAAA
KSELDKAIGRNTNGVITKDEAEKLFNQDVDAAVRGILRNAKLK
PVYDSLDAVRRAALINMVFQMGETGVAGFTNSLRMLQQKRWDE
AAVNLAKSRWYNQTPNRAKRVITTFRTGTWDAYKNL
```

Çekirdeklere neden ihtiyaç duyuyoruz?

- Belli bir yapıya sahip nesneler arasındaki benzerlik hesaplamakta kullanabiliriz.
- Örnek: İki resim arasındaki benzerlik

Çekirdeklere neden ihtiyaç duyuyoruz?

- Belli bir yapıya sahip nesneler arasındaki benzerlik hesaplamakta kullanabiliriz.
- Örnek: İki metin arasındaki benzerlik

```
(Reuters) - Developed countries face a sharp year-end slowdown led
by a contraction in Germany, the OECD said on Thursday, urging
central banks to keep rates low or pursue other forms of monetary
easing if the downturn becomes entrenched.
```

```
(Reuters) - More than 100 spacecraft have been to the moon,
including six with U.S. astronauts, but one key piece of
information about Earth's natural satellite is still missing -
what's inside.
```

Daha iyi çekirdekler tasarlayabilir miyiz?

- Evet, yapabiliriz.
  - çekirdek fonksiyonlarının parametrelerini seçerek (polinom çekirdeğinin derecesi, Gaussian çekirdeğinin yarıçapı)
  - yeni çekirdek fonksiyonları önererek
  - alan bilgisini çekirdeğin içerisinde kullanarak
  - basit çekirdekleri birleştirerek (çoklu çekirdek öğrenimi)
  - çekirdek fonksiyonunu veriden öğrenerek

Destek vektör makineleri

- Eğitim kümesi: $\{(\boldsymbol{x}_i, y_i) \in \mathcal{X} \times \{\pm 1\}\}_{i=1}^{N}$

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1D} \\ x_{21} & x_{22} & \dots & x_{2D} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \dots & x_{ND} \end{bmatrix} \qquad \boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$

- Yeni verilen bir veri noktası $\boldsymbol{x}_* \in \mathcal{X}$ için sınıf etiketini kestirmek

$$\boldsymbol{x}_* = \begin{bmatrix} x_{*1} & x_{*2} & \dots & x_{*D} \end{bmatrix} \qquad y_* = \begin{bmatrix} ? \end{bmatrix}$$

Destek vektör makineleri

- $\mathcal{X}$ uzayında bir düzlem

  $\{\boldsymbol{x} \in \mathcal{X} | \langle \boldsymbol{w}, \boldsymbol{x} \rangle + b = 0\}$

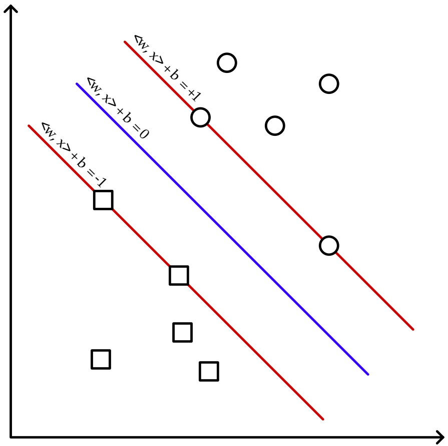
- Karar fonksiyonu: $f : \mathcal{X} \to \{\pm 1\}$

  $\boldsymbol{x} \mapsto f(\boldsymbol{x}) = \mathrm{sgn}(\langle \boldsymbol{w}, \boldsymbol{x} \rangle + b) \qquad \boldsymbol{w} \in \mathcal{X}$ ve $b \in \mathbb{R}$

- Kanonik düzlem $(\boldsymbol{w}, b)$

  $\min_{i=1,\ldots,N} |\langle \boldsymbol{w}, \boldsymbol{x}_i \rangle + b| = 1$

Destek vektör makineleri



Destek vektör makineleri

- $\|\boldsymbol{w}\|_2$'yi en küçüklersek iki sınıf arasındaki mesafeyi artırabiliriz.
- Karar fonksiyonu ve kanonik düzlem kullanılarak

  $y_i(\langle \boldsymbol{w}, \boldsymbol{x}_i \rangle + b) \geq 1 \qquad \forall i$

- Birincil en iyileme problemi

$$
\begin{aligned}
\text{en küçükle} \quad & \frac{1}{2}\|\boldsymbol{w}\|_2^2 \\
\text{karar değişkenleri} \quad & \boldsymbol{w} \in \mathcal{X}, \quad b \in \mathbb{R} \\
\text{kısıtlar} \quad & y_i(\langle \boldsymbol{w}, \boldsymbol{x}_i \rangle + b) \geq 1 \qquad \forall i
\end{aligned}
$$

Destek vektör makineleri

- Basit bir örnek

$$\begin{aligned} \text{en küçükle} \quad & w^2 - 6w + 5 \\ \text{karar değişkenleri} \quad & w \in \mathbb{R} \\ \text{kısıtlar} \quad & w - 4 \geq 0 \end{aligned}$$

- Lagrange eşiz fonksiyonu

$$L(w, \alpha) = w^2 - 6w + 5 - \alpha(w - 4)$$

- Birincil değişkene göre türev

$$\frac{\partial L(w, \alpha)}{\partial w} = 0 \Rightarrow w = \frac{\alpha + 6}{2}$$

Destek vektör makineleri

- Eşiz en iyileme problemi (w değerini Lagrange eşiz fonksiyonunun içine yazarak)

$$\begin{array}{rl} \text{en büyükle} & \dfrac{-\alpha^2 + 4\alpha - 16}{4} \\ \text{karar değişkenleri} & \alpha \in \mathbb{R}_+ \end{array}$$

- En iyi çözüm

$$\alpha^\star = 2$$

$$w^\star = \frac{\alpha^\star + 6}{2} = 4$$

Destek vektör makineleri

- Lagrange eşiz fonksiyonu

$$L(\boldsymbol{w}, b, \boldsymbol{\alpha}) = \frac{1}{2}\|\boldsymbol{w}\|_2^2 - \sum_{i=1}^{N} \alpha_i (y_i(\langle \boldsymbol{w}, \boldsymbol{x}_i \rangle + b) - 1)$$

- Birincil değişkenlere göre türev

$$\frac{\partial L(\boldsymbol{w}, b, \boldsymbol{\alpha})}{\partial \boldsymbol{w}} = 0 \Rightarrow \boldsymbol{w} = \sum_{i=1}^{N} \alpha_i y_i \boldsymbol{x}_i$$

$$\frac{\partial L(\boldsymbol{w}, b, \boldsymbol{\alpha})}{\partial b} = 0 \Rightarrow \sum_{i=1}^{N} \alpha_i y_i = 0$$

Destek vektör makineleri

- Eşiz en iyileme problemi

$$\text{en büyükle} \quad \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle$$

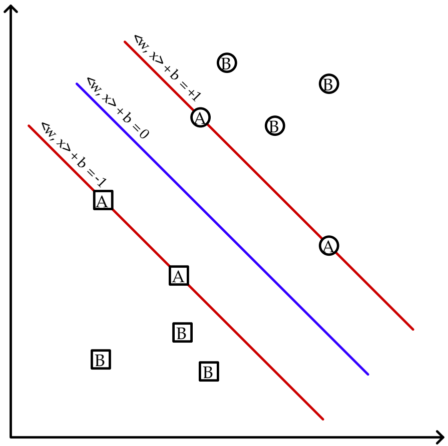
$$\text{karar değişkenleri} \quad \boldsymbol{\alpha} \in \mathbb{R}_+^N$$

$$\text{kısıtlar} \quad \sum_{i=1}^{N} \alpha_i y_i = 0$$

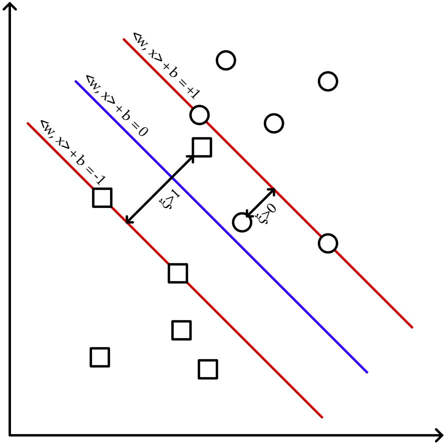
- Karar fonksiyonu

$$f(\boldsymbol{x}) = \text{sgn}\left(\sum_{i=1}^{N} \alpha_i y_i \langle \boldsymbol{x}_i, \boldsymbol{x} \rangle + b \right)$$

Destek vektör makineleri



Destek vektör makineleri



Destek vektör makineleri

- Sınıflandırma hatasına izin verdiğimiz durumda kısıtımız şu hale gelmektedir.

  $y_i(\langle \boldsymbol{w}, \boldsymbol{x}_i \rangle + b) \geq 1 - \xi_i \qquad \forall i$

- Birincil en iyileme problemi

$$
\begin{aligned}
\text{en küçükle} \quad & \frac{1}{2}\|\boldsymbol{w}\|_2^2 + C\sum_{i=1}^{N}\xi_i \\
\text{karar değişkenleri} \quad & \boldsymbol{w} \in \mathcal{X}, \quad b \in \mathbb{R}, \quad \boldsymbol{\xi} \in \mathbb{R}_+^N \\
\text{kısıtlar} \quad & y_i(\langle \boldsymbol{w}, \boldsymbol{x}_i \rangle + b) \geq 1 - \xi_i \qquad \forall i
\end{aligned}
$$

Destek vektör makineleri

- Lagrange eşiz fonksiyonu

$$L(\boldsymbol{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{1}{2}\|\boldsymbol{w}\|_2^2 + C\sum_{i=1}^{N}\xi_i - \sum_{i=1}^{N}\beta_i\xi_i - \sum_{i=1}^{N}\alpha_i(y_i(\langle \boldsymbol{w}, \boldsymbol{x}_i\rangle + b) - 1 + \xi_i)$$

- Birincil değişkenlere göre ek bir türev

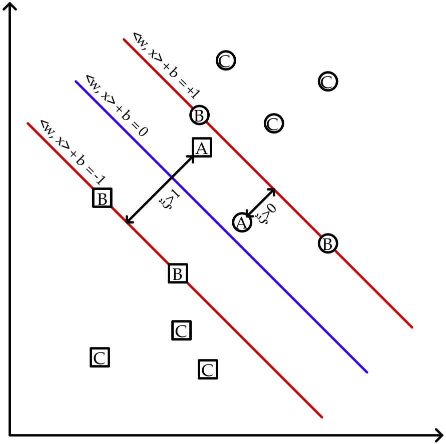
$$\frac{\partial L(\boldsymbol{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\beta})}{\partial \xi_i} = 0 \Rightarrow C = \alpha_i + \beta_i \Rightarrow C \geq \alpha_i \geq 0 \qquad \forall i$$

Destek vektör makineleri

- Eşiz en iyileme problemi

$$\begin{aligned} \text{en büyükle} \quad & \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle \\ \text{karar değişkenleri} \quad & \boldsymbol{\alpha} \in \mathbb{R}_+^N \\ \text{kısıtlar} \quad & \sum_{i=1}^{N} \alpha_i y_i = 0 \\ & \color{blue}{C \geq \alpha_i \geq 0 \qquad \forall i} \end{aligned}$$

Destek vektör makineleri



Destek vektör makineleri

- $\langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle$ terimini $k(\boldsymbol{x}_i, \boldsymbol{x}_j)$ ile değiştiriyoruz.

$$\begin{aligned} \text{en büyükle} \quad & \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j k(\boldsymbol{x}_i, \boldsymbol{x}_j) \\ \text{karar değişkenleri} \quad & \boldsymbol{\alpha} \in \mathbb{R}_+^N \\ \text{kısıtlar} \quad & \sum_{i=1}^{N} \alpha_i y_i = 0 \\ & C \geq \alpha_i \geq 0 \qquad \forall i \end{aligned}$$

- Karar fonksiyonu

$$f(\boldsymbol{x}) = \text{sgn}\left(\sum_{i=1}^{N} \alpha_i y_i k(\boldsymbol{x}_i, \boldsymbol{x}) + b \right)$$

Gauss süreci

- Eğitim kümesi: $\{(\boldsymbol{x}_i, y_i) \in \mathcal{X} \times \mathbb{R}\}_{i=1}^{N}$

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1D} \\ x_{21} & x_{22} & \cdots & x_{2D} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{ND} \end{bmatrix} \qquad \boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$

- Yeni verilen bir veri noktası $\boldsymbol{x}_* \in \mathcal{X}$ için hedef çıktıyı kestirmek

$$\boldsymbol{x}_* = \begin{bmatrix} x_{*1} & x_{*2} & \cdots & x_{*D} \end{bmatrix} \qquad y_* = \begin{bmatrix} ? \end{bmatrix}$$

Gauss süreci

- Normal dağılım: $\mathcal{N}(x; \mu, \sigma^2)$

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

- Çok değişkenli normal dağılım: $\mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$

$$p(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^D |\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}-\boldsymbol{\mu})\right)$$

Gauss süreci

$$y_i = f(\boldsymbol{x}_i) + \xi_i \qquad \forall i$$
$$f \sim \mathcal{GP}(f; m(\cdot), k(\cdot, \cdot))$$
$$\xi_i \sim \mathcal{N}(\xi_i; 0, \sigma^2) \qquad \forall i$$

$y_i:$ $\boldsymbol{x}_i$ için ölçülen çıktı değeri

$f(\boldsymbol{x}_i):$ $\boldsymbol{x}_i$ için gerçek çıktı değeri

$m(\cdot):$ ortalama fonksiyonu

$k(\cdot, \cdot):$ kovaryans fonksiyonu

$\xi_i:$ ölçüm hatası

Gauss süreci

- Gauss süreci fonksiyonlar üzerinde tanımlanmış bir dağılımdır.

- Gözlemlenen değişkenler beraberce normal dağılım izlemektedir.

$$\begin{bmatrix} f(\boldsymbol{x}_1) \\ \vdots \\ f(\boldsymbol{x}_N) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} f(\boldsymbol{x}_1) \\ \vdots \\ f(\boldsymbol{x}_N) \end{bmatrix}; \begin{bmatrix} m(\boldsymbol{x}_1) \\ \vdots \\ m(\boldsymbol{x}_N) \end{bmatrix}, \begin{bmatrix} k(\boldsymbol{x}_1, \boldsymbol{x}_1) & \ldots & k(\boldsymbol{x}_1, \boldsymbol{x}_N) \\ \vdots & \ddots & \vdots \\ k(\boldsymbol{x}_N, \boldsymbol{x}_1) & \ldots & k(\boldsymbol{x}_N, \boldsymbol{x}_N)) \end{bmatrix} \right)$$

Gauss süreci

- Ortalama fonksiyonu olarak genelde 0 fonksiyonu kullanılmaktadır.

  $m(\boldsymbol{x}_i) = 0$

- Kovaryans fonksiyonu olarak Gauss çekirdeği sıklıkla kullanılmaktadır.

  $k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp(-\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2 / s^2)$

$$\boldsymbol{y} = \boldsymbol{f} + \boldsymbol{\xi}$$
$$\boldsymbol{f} \sim \mathcal{N}(\boldsymbol{f}; \mathbf{0}, \mathbf{K})$$
$$\boldsymbol{\xi} \sim \mathcal{N}(\boldsymbol{\xi}; \mathbf{0}, \sigma^2 \mathbf{I})$$

Gauss süreci

- Bayes teoremini kullandığımızda:

$$p(\boldsymbol{f}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{f})p(\boldsymbol{f})}{p(\boldsymbol{y})}$$

$$\begin{aligned} p(\boldsymbol{f}|\boldsymbol{y}) &\propto p(\boldsymbol{y}|\boldsymbol{f})p(\boldsymbol{f}) \\ &\propto p(\boldsymbol{y}-\boldsymbol{f})p(\boldsymbol{f}) \\ &\propto \mathcal{N}(\boldsymbol{y}-\boldsymbol{f};\mathbf{0},\sigma^2\mathbf{I})\mathcal{N}(\boldsymbol{f};\mathbf{0},\mathbf{K}) \end{aligned}$$

Gauss süreci

- En büyük ardıl olasılık kestirimini kullanarak:

$$\boldsymbol{f}^{\star} = \arg\max_{\boldsymbol{f}} p(\boldsymbol{f}|\boldsymbol{y}) = \arg\max_{\boldsymbol{f}} \log p(\boldsymbol{f}|\boldsymbol{y})$$

$$\log p(\boldsymbol{f}|\boldsymbol{y}) = -\frac{1}{2\sigma^2}\|\boldsymbol{y} - \boldsymbol{f}\|_2^2 - \frac{1}{2}\boldsymbol{f}^{\top}\mathbf{K}^{-1}\boldsymbol{f} + \text{constant}$$

- $\boldsymbol{f}$ değişkenini $\mathbf{K}\boldsymbol{\alpha}$ ile yer değiştirdiğimizde:

$$\log p(\boldsymbol{\alpha}|\boldsymbol{y}) = -\frac{1}{2\sigma^2}\|\boldsymbol{y} - \mathbf{K}\boldsymbol{\alpha}\|_2^2 - \frac{1}{2}\boldsymbol{\alpha}^{\top}\mathbf{K}\boldsymbol{\alpha} + \text{constant}$$

Gauss süreci

- Kısıtsız en iyileme problemi

$$\begin{aligned}\boldsymbol{\alpha}^\star &= \arg\max_{\boldsymbol{\alpha}} \log p(\boldsymbol{\alpha}|\boldsymbol{y}) \\ &= \arg\max_{\boldsymbol{\alpha}} -\frac{1}{2\sigma^2}\left(\boldsymbol{y}^\top\boldsymbol{y} - 2\boldsymbol{y}^\top\mathbf{K}\boldsymbol{\alpha} + \boldsymbol{\alpha}^\top\mathbf{K}\mathbf{K}\boldsymbol{\alpha}\right) - \frac{1}{2}\boldsymbol{\alpha}^\top\mathbf{K}\boldsymbol{\alpha}\end{aligned}$$

- α değişkenine göre türev alıp en iyi değeri bulabiliriz.

$$\boldsymbol{\alpha}^\star = (\mathbf{K} + \sigma^2\mathbf{I})^{-1}\boldsymbol{y}$$

Gauss süreci

$$\boldsymbol{a} \sim \mathcal{N}(\boldsymbol{a}; \boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a)$$
$$\boldsymbol{b} \sim \mathcal{N}(\boldsymbol{b}; \boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b)$$
$$\boldsymbol{a} + \boldsymbol{b} \sim \mathcal{N}(\boldsymbol{a} + \boldsymbol{b}; \boldsymbol{\mu}_a + \boldsymbol{\mu}_b, \boldsymbol{\Sigma}_a + \boldsymbol{\Sigma}_b)$$

- $\boldsymbol{y}$ and y_* için bileşik olasılık dağılımı

$$\begin{bmatrix} \boldsymbol{y} \\ y_* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{y} \\ y_* \end{bmatrix}; \begin{bmatrix} \mathbf{0} \\ 0 \end{bmatrix}, \begin{bmatrix} \mathbf{K} + \sigma^2 \mathbf{I} & \boldsymbol{k}_* \\ \boldsymbol{k}_*^\top & k(\boldsymbol{x}_*, \boldsymbol{x}_*) + \sigma^2 \end{bmatrix} \right)$$

$$\boldsymbol{k}_* = \begin{bmatrix} k(\boldsymbol{x}_1, \boldsymbol{x}_*) \\ \vdots \\ k(\boldsymbol{x}_N, \boldsymbol{x}_*) \end{bmatrix}$$

Gauss süreci

$$\begin{bmatrix} \boldsymbol{a} \\ \boldsymbol{b} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{a} \\ \boldsymbol{b} \end{bmatrix}; \begin{bmatrix} \boldsymbol{\mu}_a \\ \boldsymbol{\mu}_b \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{ab} \\ \boldsymbol{\Sigma}_{ba} & \boldsymbol{\Sigma}_{bb} \end{bmatrix} \right)$$

$$\boldsymbol{b}|\boldsymbol{a} \sim \mathcal{N}\big(\boldsymbol{b}; \boldsymbol{\mu}_b + \boldsymbol{\Sigma}_{ba}\boldsymbol{\Sigma}_{aa}^{-1}(\boldsymbol{a} - \boldsymbol{\mu}_a), \boldsymbol{\Sigma}_{bb} - \boldsymbol{\Sigma}_{ba}\boldsymbol{\Sigma}_{aa}^{-1}\boldsymbol{\Sigma}_{ab}\big)$$

- y_* değerini kestirmek için aşağıdaki dağılımı kullanabiliriz.

$$y_*|\boldsymbol{y} \sim \mathcal{N}\Big(y_*; \boldsymbol{k}_*^\top \underbrace{(\mathbf{K} + \sigma^2\mathbf{I})^{-1}\boldsymbol{y}}_{\boldsymbol{\alpha}}, k(\boldsymbol{x}_*, \boldsymbol{x}_*) + \sigma^2 - \boldsymbol{k}_*^\top(\mathbf{K} + \sigma^2\mathbf{I})^{-1}\boldsymbol{k}_*\Big)$$

Çekirdek tasarımı

- Destek vektör makinesinin eşiz en iyileme problemi dual standart halde şu şekilde yazılabilir.

$$
\begin{aligned}
\text{en küçükle} \quad & -\mathbf{1}^\top \boldsymbol{\alpha} + \frac{1}{2}\boldsymbol{\alpha}^\top ((\boldsymbol{y}\boldsymbol{y}^\top) \odot \mathbf{K})\boldsymbol{\alpha} \\
\text{karar değişkenleri} \quad & \boldsymbol{\alpha} \in \mathbb{R}^N \\
\text{kısıtlar} \quad & \boldsymbol{y}^\top \boldsymbol{\alpha} = 0 \\
& \mathbf{0} \leq \boldsymbol{\alpha} \leq C\mathbf{1}
\end{aligned}
$$

- En iyilemenin uygun bir şekilde gerçekleşebilmesi için $\mathbf{K}$ matrisinin yarı-kesin pozitif matris olması gerekmektedir.

Çekirdek tasarımı

- $\mathbb{R}^{N\times N}$ kümesinden alınan simetrik bir matris $\mathbf{K}$ eğer bütün özdeğerleri sıfırdan büyük eşit ise yarı-kesin pozitiftir.

$$\boldsymbol{\alpha}^\top \mathbf{K}\boldsymbol{\alpha} \geq 0 \qquad \forall \boldsymbol{\alpha} \in \mathbb{R}^N.$$

- Bütün geçerli çekirdek fonksiyonları yarı-kesin pozitif çekirdek matrisleri oluşturmaktadır.

$$\begin{aligned}\boldsymbol{\alpha}^\top \mathbf{K}\boldsymbol{\alpha} &= \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j k(\boldsymbol{x_i}, \boldsymbol{x_j}) = \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j \langle \Phi(\boldsymbol{x_i}), \Phi(\boldsymbol{x_j})\rangle \\ &= \left\langle \sum_{i=1}^{N} \alpha_i \Phi(\boldsymbol{x_i}), \sum_{j=1}^{N} \alpha_j \Phi(\boldsymbol{x_j}) \right\rangle = \left\| \sum_{i=1}^{N} \alpha_i \Phi(\boldsymbol{x_i}) \right\|_2^2 \geq 0\end{aligned}$$

Çekirdek tasarımı

- k_1 ve k_2 $\mathcal{X} \times \mathcal{X}$ üzerinde tanımlanmış çekirdek fonksiyonları, $\mathcal{X} \subseteq \mathbb{R}^D$, $a \in \mathbb{R}_+$, $f(\cdot)$ $\mathcal{X}$ üzerinde tanımlanmış gerçel değerler üreten bir fonksiyon, k_3 ise $\mathbb{R}^S \times \mathbb{R}^S$ üzerinde tanımlanmış bir çekirdek fonksiyonu ($\Phi \colon \mathcal{X} \rightarrow \mathbb{R}^S$) ve $\mathbf{L}$ $D \times D$ boyutunda yarı-kesin pozitif bir çekirdek matrisi olmak üzere aşağıdaki fonksiyonlar da geçerli çekirdek fonksiyonlarıdır.
  - i $k(\boldsymbol{x}_i, \boldsymbol{x}_j) = k_1(\boldsymbol{x}_i, \boldsymbol{x}_j) + k_2(\boldsymbol{x}_i, \boldsymbol{x}_j)$
  - ii $k(\boldsymbol{x}_i, \boldsymbol{x}_j) = a k_1(\boldsymbol{x}_i, \boldsymbol{x}_j)$
  - iii $k(\boldsymbol{x}_i, \boldsymbol{x}_j) = k_1(\boldsymbol{x}_i, \boldsymbol{x}_j) k_2(\boldsymbol{x}_i, \boldsymbol{x}_j)$
  - iv $k(\boldsymbol{x}_i, \boldsymbol{x}_j) = f(\boldsymbol{x}_i) f(\boldsymbol{x}_j)$
  - v $k(\boldsymbol{x}_i, \boldsymbol{x}_j) = k_3(\Phi(\boldsymbol{x}_i), \Phi(\boldsymbol{x}_j))$
  - vi $k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{x}_i^\top \mathbf{L} \boldsymbol{x}_j$

Çekirdek tasarımı

- k_1 $\mathcal{X} \times \mathcal{X}$ üzerinde tanımlanmış çekirdek fonksiyonu ve p $\mathcal{X}$ üzerinde tanımlanmış katsayıları pozitif olan bir polinom olmak üzere aşağıdaki fonksiyonlar da geçerli çekirdek fonksiyonlarıdır.

  vii $k(\boldsymbol{x}_i, \boldsymbol{x}_j) = p(k_1(\boldsymbol{x}_i, \boldsymbol{x}_j))$

  viii $k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp(k_1(\boldsymbol{x}_i, \boldsymbol{x}_j))$

  ix $k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp(-\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2/(2s^2))$

Çekirdek tasarımı

- Tek bir çekirdek fonksiyonu kullanmak yerine çok sayıda çekirdek fonksiyonu kullanabiliriz.

  $$k_\eta(\boldsymbol{x}_i, \boldsymbol{x}_j) = f_\eta(\{k_m(\boldsymbol{x}_i^m, \boldsymbol{x}_j^m)\}_{m=1}^P)$$

  Birleştirme fonksiyonu $f_\eta \colon \mathbb{R}^P \to \mathbb{R}$ çekirdek fonskiyonlarını ($\{k_m(\boldsymbol{x}_i^m, \boldsymbol{x}_j^m)\}_{m=1}^P$) doğrusal ya da doğrusal olmayan bir yöntemle birleştirebilir.

Çekirdek tasarımı

- Sabit kurallar

$$k_\eta(\boldsymbol{x}_i, \boldsymbol{x}_j) = \prod_{m=1}^{P} k_m(\boldsymbol{x}_i^m, \boldsymbol{x}_j^m)$$

$$k_\eta(\boldsymbol{x}_i, \boldsymbol{x}_j) = \sum_{m=1}^{P} k_m(\boldsymbol{x}_i^m, \boldsymbol{x}_j^m)$$

$$\langle \Phi_\eta(\boldsymbol{x}_i), \Phi_\eta(\boldsymbol{x}_j) \rangle = \begin{pmatrix} \Phi_1(\boldsymbol{x}_i^1) \\ \Phi_2(\boldsymbol{x}_i^2) \\ \vdots \\ \Phi_P(\boldsymbol{x}_i^P) \end{pmatrix}^\top \begin{pmatrix} \Phi_1(\boldsymbol{x}_j^1) \\ \Phi_2(\boldsymbol{x}_j^2) \\ \vdots \\ \Phi_P(\boldsymbol{x}_j^P) \end{pmatrix} = \sum_{m=1}^{P} k_m(\boldsymbol{x}_i^m, \boldsymbol{x}_j^m)$$

Çekirdek tasarımı

- Toplam kuralını parametrik bir hale getirebiliriz.

$$k_\eta(\boldsymbol{x}_i, \boldsymbol{x}_j) = \sum_{m=1}^{P} \eta_m k_m(\boldsymbol{x}_i^m, \boldsymbol{x}_j^m)$$

$$\langle \Phi_\eta(\boldsymbol{x}_i), \Phi_\eta(\boldsymbol{x}_j) \rangle = \begin{pmatrix} \sqrt{\eta_1}\Phi_1(\boldsymbol{x}_i^1) \\ \sqrt{\eta_2}\Phi_2(\boldsymbol{x}_i^2) \\ \vdots \\ \sqrt{\eta_P}\Phi_P(\boldsymbol{x}_i^P) \end{pmatrix}^\top \begin{pmatrix} \sqrt{\eta_1}\Phi_1(\boldsymbol{x}_j^1) \\ \sqrt{\eta_2}\Phi_2(\boldsymbol{x}_j^2) \\ \vdots \\ \sqrt{\eta_P}\Phi_P(\boldsymbol{x}_j^P) \end{pmatrix} = \sum_{m=1}^{P} \eta_m k_m(\boldsymbol{x}_i^m, \boldsymbol{x}_j^m)$$

- Doğrusal toplam ($\boldsymbol{\eta} \in \mathbb{R}^P$), konik toplam ($\boldsymbol{\eta} \in \mathbb{R}_+^P$) ve dışbükey toplam ($\boldsymbol{\eta} \in \mathbb{R}_+^P$ ve $\mathbf{1}^\top \boldsymbol{\eta} = 1$)

Çekirdek tasarımı

- Çekirdek matrisleri $\mathbf{K}_1$ ve $\mathbf{K}_2$ arasındaki hizalanma miktarı şu şekilde yazılır.

$$A(\mathbf{K}_1, \mathbf{K}_2) = \frac{\langle \mathbf{K}_1, \mathbf{K}_2 \rangle_F}{\sqrt{\langle \mathbf{K}_1, \mathbf{K}_1 \rangle_F \langle \mathbf{K}_2, \mathbf{K}_2 \rangle_F}}$$

$$\langle \mathbf{K}_1, \mathbf{K}_2 \rangle_F = \sum_{i=1}^{N} \sum_{j=1}^{N} k_1(\boldsymbol{x}_i, \boldsymbol{x}_j) k_2(\boldsymbol{x}_i, \boldsymbol{x}_j).$$

- Bu değer iki matrisin N^2 boyutunda vektörlere çevrilip aralarındaki açının kosinüsünü hesaplamaya karşılık gelmektedir ($-1 \leq A(\mathbf{K}_1, \mathbf{K}_2) \leq +1$).

Çekirdek tasarımı

- İkili sınıflandırma için ideal çekirdek fonksiyonu:

$$\boldsymbol{y}\boldsymbol{y}^\top = \begin{bmatrix} y_1 y_1 & y_1 y_2 & \cdots & y_1 y_N \\ y_2 y_1 & y_2 y_2 & \cdots & y_2 y_N \\ \vdots & \vdots & \ddots & \vdots \\ y_N y_1 & y_N y_2 & \cdots & y_N y_N \end{bmatrix}$$

- Bu ideal çekirdeğe benzeyen bir çekirdek oluşturmaya çalışabiliriz.

$$\eta_m = \frac{A(\mathbf{K}_m, \boldsymbol{y}\boldsymbol{y}^\top)}{\sum_{h=1}^{P} A(\mathbf{K}_h, \boldsymbol{y}\boldsymbol{y}^\top)} \qquad \forall m$$

Çekirdek tasarımı

- Çekirdek ağırlıklarını birleştirilmiş çekirdekle ideal çekirdek arasındaki hizalanmayı en büyükleyerek seçebiliriz.

$$A(\mathbf{K}_\eta, \boldsymbol{y}\boldsymbol{y}^\top) = \frac{\sum_{m=1}^{P} \eta_m \langle \mathbf{K}_m, \boldsymbol{y}\boldsymbol{y}^\top \rangle_F}{N\sqrt{\sum_{m=1}^{P} \sum_{h=1}^{P} \eta_m \eta_h \langle \mathbf{K}_m, \mathbf{K}_h \rangle_F}}$$

$$\begin{aligned} \text{en büyükle} \quad & \sum_{m=1}^{P} \eta_m \langle \mathbf{K}_m, \boldsymbol{y}\boldsymbol{y}^\top \rangle_F \\ \text{karar değişkenleri} \quad & \boldsymbol{\eta} \in \mathbb{R}_+^P \\ \text{kısıtlar} \quad & \sum_{m=1}^{P} \sum_{h=1}^{P} \eta_m \eta_h \langle \mathbf{K}_m, \mathbf{K}_h \rangle_F = c \end{aligned}$$

Çekirdek tasarımı

- Çekirdek ağırlıklarını birleştirilmiş çekirdekle ideal çekirdek arasındaki mesafeyi en küçükleyerek de seçebiliriz.

$$\begin{aligned}\text{en küçükle} \quad & \langle \mathbf{K}_\eta - \boldsymbol{y}\boldsymbol{y}^\top, \mathbf{K}_\eta - \boldsymbol{y}\boldsymbol{y}^\top \rangle_F \\ \text{karar değişkenleri} \quad & \boldsymbol{\eta} \in \mathbb{R}_+^P \\ \text{kısıtlar} \quad & \sum_{m=1}^{P} \eta_m = 1\end{aligned}$$

$$\begin{aligned}\text{en küçükle} \quad & \sum_{m=1}^{P} \sum_{h=1}^{P} \eta_m \eta_h \langle \mathbf{K}_m, \mathbf{K}_h \rangle_F - 2 \sum_{m=1}^{P} \eta_m \langle \mathbf{K}_m, \boldsymbol{y}\boldsymbol{y}^\top \rangle_F \\ \text{karar değişkenleri} \quad & \boldsymbol{\eta} \in \mathbb{R}_+^P \\ \text{kısıtlar} \quad & \sum_{m=1}^{P} \eta_m = 1\end{aligned}$$