En Küçük Kareler

T.C. Trakya Üniversitesi
Mühendislik Fakültesi
Elektrik - Elektronik Mühendisliği Bölümü
Kontrol Anabilim Dalı

Dr. Öğr. Üyesi Işık İlber Sırmatel
sirmatel.github.io

Kaynak (source)

Lecture Slides for Introduction to Applied Linear Algebra: Vectors, Matrices, and Least Squares.
Stephen Boyd, Lieven Vandenberghe

Konu listesi

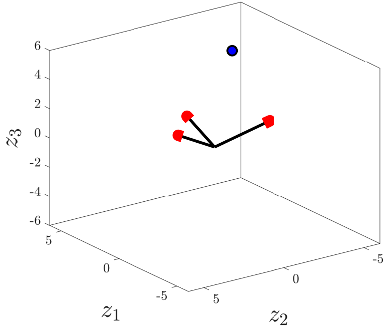
Bölüm 1

Doğrusal denklemlerin geometrisi

Doğrusal denklemlerin geometrisi, örnek 1

$$Ax = b \qquad A = \begin{bmatrix} 1 & -1 & -2 \\ 1 & 3 & -2 \\ 2 & 2 & 2 \end{bmatrix} \qquad b = \begin{bmatrix} 4 \\ -4 \\ 4 \end{bmatrix}$$

$$x = A^{-1}b = \begin{bmatrix} 0.417 & -0.083 & 0.333 \\ -0.25 & 0.25 & 0 \\ -0.167 & -0.167 & 0.167 \end{bmatrix} \begin{bmatrix} 4 \\ -4 \\ 4 \end{bmatrix} = \begin{bmatrix} 3.333 \\ -2 \\ 0.667 \end{bmatrix}$$

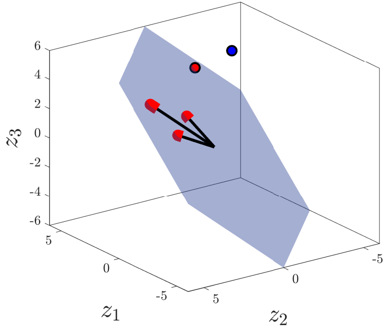


- kare denklem takımı (3 denklem, 3 bilinmeyen)
- A^{-1} mevcut
- $b \in \mathcal{R}(A)$
- $Ax = b$'nin eşsiz bir çözümü mevcut

Doğrusal denklemlerin geometrisi, örnek 2

$$Ax = b \qquad A = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 3 & 4 \\ 2 & 2 & 4 \end{bmatrix} \qquad b = \begin{bmatrix} 4 \\ -4 \\ 4 \end{bmatrix}$$

$$\hat{x}_{\mathsf{LS}} = A^{\dagger} b = \begin{bmatrix} 0.36 & -0.14 & 0.22 \\ -0.31 & 0.19 & -0.11 \\ 0.06 & 0.06 & 0.11 \end{bmatrix} \begin{bmatrix} 4 \\ -4 \\ 4 \end{bmatrix} = \begin{bmatrix} 2.889 \\ -2.444 \\ 0.444 \end{bmatrix}$$

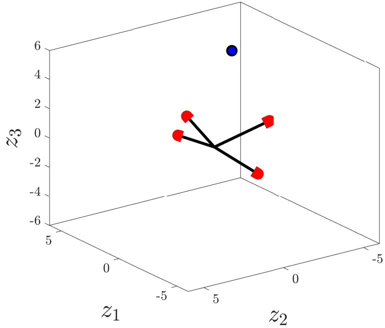


- kare denklem takımı (3 denklem, 3 bilinmeyen)
- A^{-1} mevcut değil
- $b \notin \mathcal{R}(A)$
- $Ax = b$'nin çözümü mevcut değil

Doğrusal denklemlerin geometrisi, örnek 3

$$Ax = b \qquad A = \begin{bmatrix} 1 & -1 & -2 & -1 \\ 1 & 3 & -2 & -2 \\ 2 & 2 & 2 & -2 \end{bmatrix} \qquad b = \begin{bmatrix} 4 \\ -4 \\ 4 \end{bmatrix}$$

$$\hat{x}_{\mathsf{LN}} = A^{\dagger} b = \begin{bmatrix} 0.25 & -0.09 & 0.2 \\ -0.29 & 0.25 & -0.04 \\ -0.14 & -0.17 & 0.19 \\ -0.18 & -0.01 & -0.14 \end{bmatrix} \begin{bmatrix} 4 \\ -4 \\ 4 \end{bmatrix} = \begin{bmatrix} 2.17 \\ -2.32 \\ 0.88 \\ -1.27 \end{bmatrix}$$

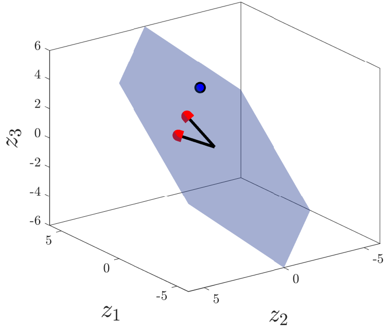


- eksik-belirli denklem takımı (3 denklem, 4 bilinmeyen)
- A^{-1} mevcut değil
- $b \in \mathcal{R}(A)$
- $Ax = b$'nin sonsuz adet çözümü mevcut

Doğrusal denklemlerin geometrisi, örnek 4

$$Ax = b \qquad A = \begin{bmatrix} 1 & -1 \\ 1 & 3 \\ 2 & 2 \end{bmatrix} \qquad b = \begin{bmatrix} 4 \\ -2 \\ 2 \end{bmatrix}$$

$$x = A^{\dagger}b = \begin{bmatrix} 0.417 & -0.083 & 0.333 \\ -0.25 & 0.25 & 0 \end{bmatrix} \begin{bmatrix} 4 \\ -2 \\ 2 \end{bmatrix} = \begin{bmatrix} 2.5 \\ -1.5 \end{bmatrix}$$

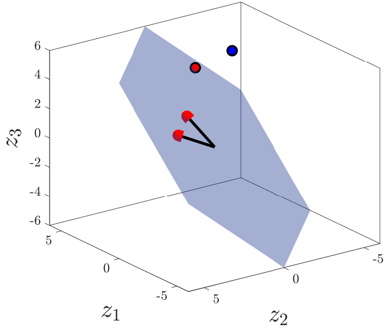


- aşırı-belirli denklem takımı (3 denklem, 2 bilinmeyen)
- A^{-1} mevcut değil
- $b \in \mathcal{R}(A)$
- $Ax = b$'nin eşsiz bir çözümü mevcut

Doğrusal denklemlerin geometrisi, örnek 5

$$Ax = b \qquad A = \begin{bmatrix} 1 & -1 \\ 1 & 3 \\ 2 & 2 \end{bmatrix} \qquad b = \begin{bmatrix} 4 \\ -4 \\ 4 \end{bmatrix}$$

$$\hat{x}_{LS} = A^{\dagger} b = \begin{bmatrix} 0.417 & -0.083 & 0.333 \\ -0.25 & 0.25 & 0 \end{bmatrix} \begin{bmatrix} 4 \\ -4 \\ 4 \end{bmatrix} = \begin{bmatrix} 3.333 \\ -2 \end{bmatrix}$$



- aşırı-belirli denklem takımı (3 denklem, 2 bilinmeyen)
- A^{-1} mevcut değil
- $b \notin \mathcal{R}(A)$
- $Ax = b$'nin çözümü mevcut değil

Bölüm 2

En küçük kareler problemi

En küçük kareler problemi

- $m \times n$ A matrisi uzun matris olsun, bu durumda $Ax = b$ denklemi aşırı-belirli olur
- çoğu b için $Ax = b$'yı sağlayan x yoktur
- en küçük kareler (*least squares*, LS) problemi: $\|Ax - b\|^2$'yi minimize eden x'i seç
- $Ax - b$ terimine kalıntı (*residual*) denir ve r ile gösterilir: $r = Ax - b$

En küçük kareler problemi

- $\|Ax - b\|^2$ terimine amaç fonksiyonu (*objective function*) denir
- herhangi bir n-vektör x için

$$\|A\hat{x} - b\|^2 \leq \|Ax - b\|^2$$

  ise $\hat{x}$ en küçük kareler probleminin bir çözümüdür
- fikir: $\hat{x}$ kalıntının (0 olmasa da) mümkün olan en küçük değeri almasını sağlar
- en küçük kareler (veri uydurmada) bağlanım (*regression*) da denir

En küçük kareler problemi

- $\hat{x}$'e $Ax = b$'nin en küçük kareler yaklaşık çözümü (*least squares approximate solution*) denir
- $\hat{x}$'e bazen en küçük kareler anlamında $Ax = b$'nin çözümü denir ancak bu yanlıştır çünkü $\hat{x}$ $Ax = b$'nin çözümü değildir
- $\hat{x}$'nin $A\hat{x} = b$'i sağlaması gerekmez, ve genellikle de sağlamaz
- ancak $\hat{x}$ $A\hat{x} = b$'i sağlarsa en küçük kareler probleminin çözümüdür

Sütun yorumu

- $a_1, a_2, \ldots, a_n$ A'nın sütunları olsun
- en küçük kareler amaç fonksiyonu

$$\|Ax - b\|^2 = \|(x_1 a_1 + x_2 a_2 + \cdots + x_n a_n) - b\|^2$$

  şeklindedir
- dolayısıyla, en küçük kareler problemi, A'nın sütunlarının b'ye en yakın doğrusal bileşimini bulma problemidir
- $\hat{x}$ en küçük kareler probleminin bir çözümü ise

$$A\hat{x} = \hat{x}_1 a_1 + \hat{x}_2 a_2 + \cdots + \hat{x}_n a_n$$

  olarak verilen m-vektör, A'nın sütunlarının bütün doğrusal bileşimleri arasında b'ye en yakın olandır

Satır yorumu

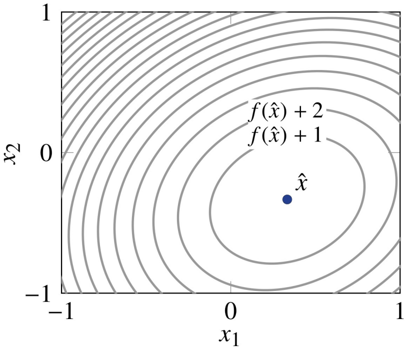
- $\tilde{a}_1^T, \tilde{a}_2^T, \ldots, \tilde{a}_n^T$ A'nın satırları olsun
- kalıntı bileşenleri (yani, kalıntı vektörü r'nin elemanları, kısaca: kalıntılar) $r_i = \tilde{a}_i^T x - b_i$ olarak yazılır
- en küçük kareler amaç fonksiyonu

$$\|Ax - b\|^2 = \sum_{i=1}^{m} r_i^2 = (\tilde{a}_1^T x - b_1)^2 + \cdots + (\tilde{a}_m^T x - b_m)^2$$

  (yani, kalıntıların karelerinin toplamı) şeklindedir
- dolayısıyla, en küçük kareler problemi, kalıntıların karelerinin toplamını minimize etme problemidir
  - $Ax = b$'yi çözmek bütün kalıntıları 0 yapmaktır
  - aşırı belirli olduğu için $Ax = b$'yi çözmek mümkün olmadığında en küçük kareler ile kalıntıların hepsinin küçük olmasına çalışılır

Örnek

$$A = \begin{bmatrix} 2 & 0 \\ -1 & 1 \\ 0 & 2 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$$



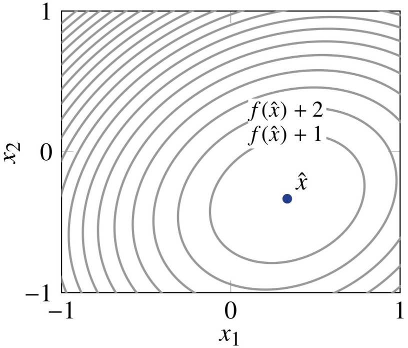
- $Ax = b$'nin çözümü yok
  - en küçük kareler problemi

$$\|Ax - b\|^2 = (2x_1 - 1)^2 + (-x_1 + x_2)^2 + (2x_2 + 1)^2$$

ifadesini minimize edecek x'i seçme problemidir

Örnek

$$A = \begin{bmatrix} 2 & 0 \\ -1 & 1 \\ 0 & 2 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$$



- en küçük kareler yaklaşık çözüm $\hat{x} = \begin{bmatrix} \frac{1}{3} & -\frac{1}{3} \end{bmatrix}^T$ şeklindedir (burada çözüm diferansiyel hesap (*calculus*) ile (yani, $\nabla \|Ax - b\|^2 = 0$'i çözerek) bulunabilir)
- $\|A\hat{x} - b\|^2 = \frac{2}{3}$, $\|Ax - b\|^2$'nin mümkün olan en küçük değeridir
- $A\hat{x} = \begin{bmatrix} \frac{2}{3} & -\frac{2}{3} & -\frac{2}{3} \end{bmatrix}^T$ A'nın sütunlarının bütün doğrusal bileşimleri arasında b'ye en yakın olandır

Bölüm 3

En küçük kareler probleminin çözümü

En küçük kareler probleminin çözümü

- bir varsayımda bulunuyoruz: A'nın sütunları doğrusal bağımsız
- dolayısıyla, A'nın Gram matrisi A^TA tersi alınabilirdir
- en küçük kareler probleminin eşsiz çözümü

$$\hat{x} = (A^TA)^{-1}A^Tb = A^\dagger b$$

- karşılaştırın: $x = A^{-1}b$ (tersi alınabilir A için $Ax = b$'nin çözümü)

Diferansiyel hesap ile türetme

en küçük kareler probleminin çözümünü diferansiyel hesap (gradyanı alıp 0'a eşitleme) kullanarak türetelim

- amaç fonksiyonunu $f(x)$ olarak tanımlayalım:

$$f(x) = \|Ax - b\|^2 = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} A_{ij} x_j - b_i \right)^2$$

- çözüm $\hat{x}$, $\nabla f(\hat{x}) = 0$'ı sağlar
- $f(x) = \|Ax - b\|^2 = (Ax - b)^T (Ax - b)$
- $f(x) = x^T A^T A x - 2b^T A x + b^T b$
- $\nabla f(\hat{x}) = 2A^T (A\hat{x} - b) = 0$
- $\hat{x}$, $(A^T A)\hat{x} = A^T b$'yi sağlar (buna, normal denklem (*normal equation*) denir)
- dolayısıyla: $\hat{x} = (A^T A)^{-1} A^T b$

Doğrudan teyit etme

- $\hat{x}$ en küçük kareler probleminin çözümü olsun: $\hat{x} = (A^T A)^{-1} A^T b$
- $\hat{x}$ çözüm olduğundan normal denklemi sağlar: $A^T(A\hat{x} - b) = 0$
- herhangi bir n-vektör x için:

$$
\begin{aligned}
\|Ax - b\|^2 &= \|(Ax - A\hat{x}) + (A\hat{x} - b)\|^2 \\
&= \|A(x - \hat{x})\|^2 + \|A\hat{x} - b\| + 2(A(x - \hat{x}))^T (A\hat{x} - b) \\
&= \|A(x - \hat{x})\|^2 + \|A\hat{x} - b\| + 2(x - \hat{x})^T \underbrace{A^T(A\hat{x} - b)}_{=0} \\
&= \|A(x - \hat{x})\|^2 + \|A\hat{x} - b\|^2
\end{aligned}
$$

- dolayısıyla, herhangi bir x için: $\|Ax - b\|^2 \geq \|A\hat{x} - b\|^2$
- $\|Ax - b\|^2 = \|A\hat{x} - b\|^2$ olursa $A(x - \hat{x}) = 0$ olur, bu da $x = \hat{x}$'i gerektirir (A'nın sütunları doğrusal bağımsız olduğundan, sadece $z = 0$ için $Az = 0$ olur)

LS yaklaşık çözümlerinin hesaplanması

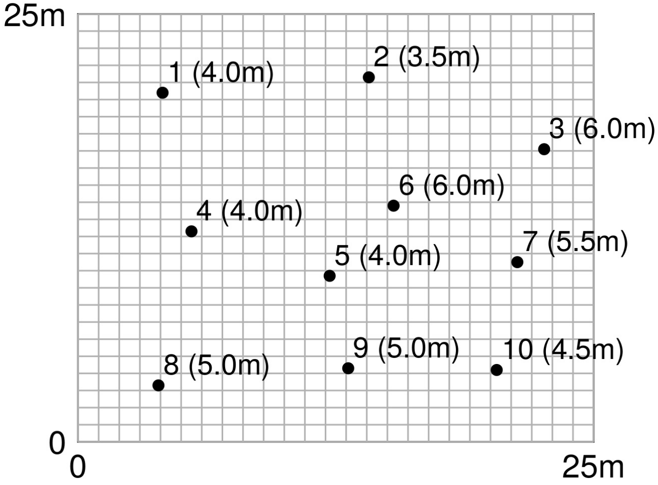
- A'nın QR ayrıştırmasını hesapla: $A = QR$ ($2mn^2$ flop)
- A'nın sütunları doğrusal bağımsız olduğundan QR ayrıştırması mevcuttur
- $\hat{x} = A^\dagger b = R^{-1}Q^T b$'yi hesaplamak için:
  - $Q^T b$'yi oluştur ($2mn$ flop)
  - geri yönde yerine koyma ile $\hat{x} = R^{-1}(Q^T b)$'yi hesapla (n^2 flop)
- toplam maliyet: yaklaşık $2mn^2$ flop
- bu, tersi alınabilir A için $Ax = b$'nin çözümünde kullanılan algoritmayla aynıdır
- ancak, A uzun matris olduğunda, algoritma en küçük kareler yaklaşık çözümü verir

Bölüm 4

Uygulama örnekleri

Aydınlatma

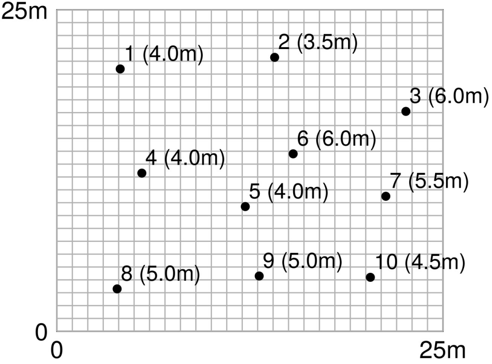
amaç: m bölgeye ayrılmış bir alanı istenen şekilde aydınlatmak için, konumları sabit n adet lambanın güçlerinin belirlenmesi



Aydınlatma

- sadece lamba j açık olduğunda (gücü 1, diğer lambalar kapalı) bölge i'deki aydınlatmayı A_{ij} ile gösterelim
- lamba j'nin gücünü x_j ile gösterelim
- bölge i'deki toplam aydınlatma $(Ax)_i$ olur
- bölge i için aydınlatma hedefini b_i ile gösterelim

Aydınlatma

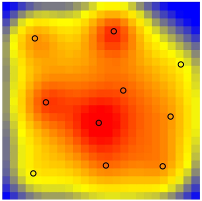


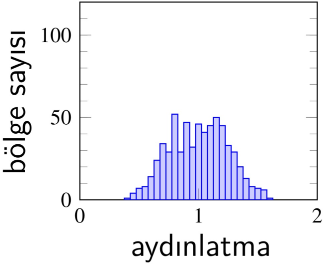
örnekteki alan için:
$m = 25^2 = 625$, $n = 10$
(dolayısıyla $Ax = b$'de 625 denklem ve 10 bilinmeyen var)

yandaki şekildeki noktalar lambaların konumlarını göstermektedir

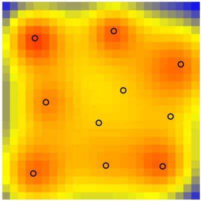
Aydınlatma

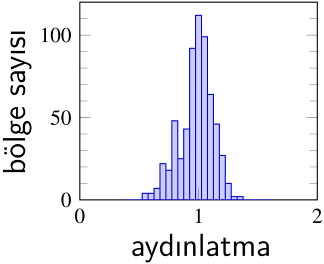
- lambalar eşit güçte ($x = \mathbf{1}$)





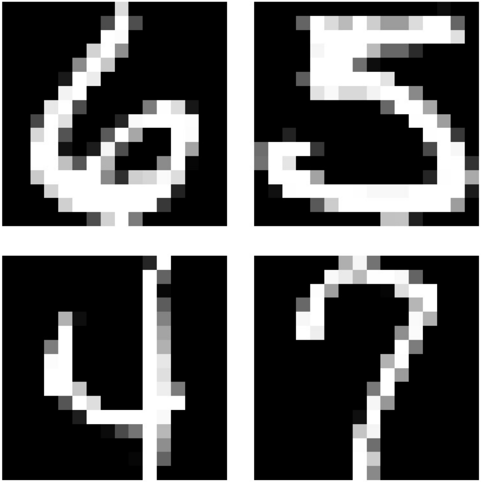
- en küçük kareler çözümü $\hat{x}$ ($b = \mathbf{1}$ için)





El yazısı rakam sınıflandırma

amaç: el yazısı rakam içeren bir görüntünün hangi rakamı gösterdiğini otomatik olarak belirlemek



El yazısı rakam sınıflandırma

- görüntüler (*image*) 16×16 piksel, 256-vektörler olarak temsil ediliyorlar
- değerler $[0, 1]$ aralığında (0 siyah, 1 beyaz)
- örneğin, "görüntüdeki rakam 0 mı?" diye sınıflandırma (*classification*) yapalım
- etiketler (*label*): görüntü 0 ise $y_i = 1$, değilse $y_i = -1$
- benzer bir veritabanı: MNIST database, ilgili bir yayın: LeCun, Yann, Léon Bottou, Yoshua Bengio, and Patrick Haffner. "Gradient-based learning applied to document recognition." Proceedings of the IEEE 86, no. 11 (1998): 2278-2324.

El yazısı rakam sınıflandırma

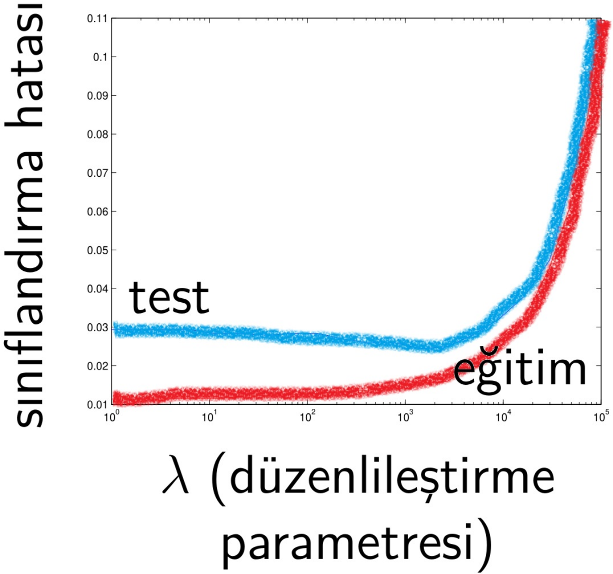
- Boole sınıflandırıcı (*Boolean classifier*):
  $\hat{y} = \mathrm{sign}(w^T x + v)$
- 256-vektör w ağırlık (*weight*), v kayma (*offset*)
- $y_i \approx \hat{y}_i = \mathrm{sign}(w^T x_i + v)$'yi sağlayan w ve v'yi bulmak istiyoruz
- eğitim veri kümesi (*training set*): (x_i, y_i) çiftleri (N adet)
- bu veri kümesi ile en küçük kareler problemi çözülerek

  $$\sum_{i=1}^{N}(w^T x_i + v - y_i)^2 + \lambda\|w\|^2$$

  ifadesini minimize eden w ve v seçilir
- test veri kümesi (*test set*) üzerinde sınıflandırıcı test edilir
- buradaki λ düzenlileştirme (*regularization*) parametresidir

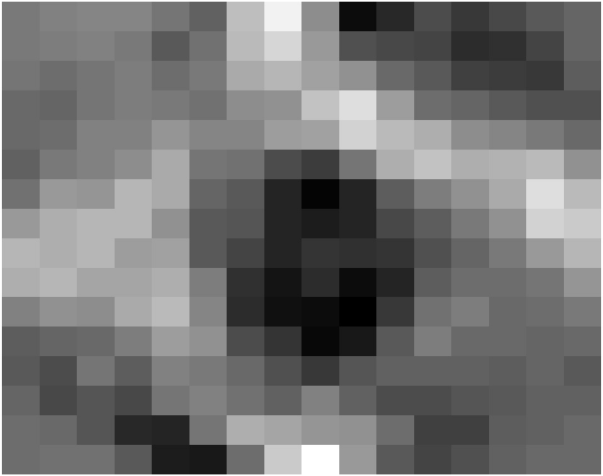
El yazısı rakam sınıflandırma

el yazısı rakam 0'ı tahmin ederken yapılan sınıflandırma hatası



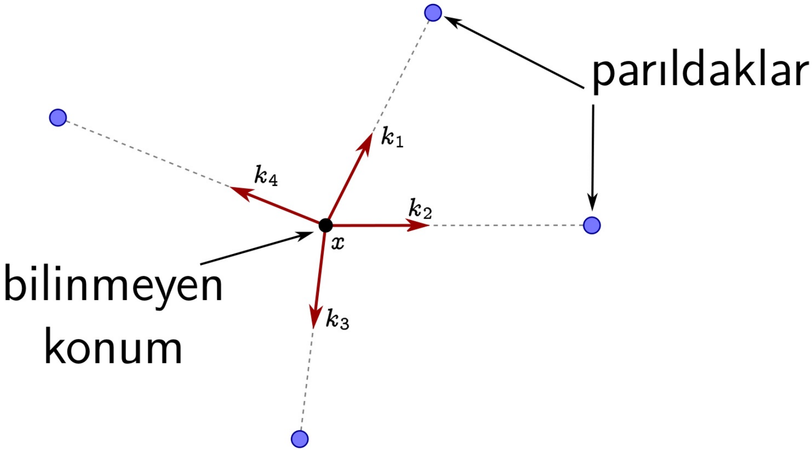
El yazısı rakam sınıflandırma

ağırlık vektörü w (en küçük kareler probleminin çözümü)



Erim ölçümleriyle yöngüdüm

amaç: m adet parıldaktan (*beacon*) gelen erim ölçümleri (*range measurements*) ile bilinmeyen bir n boyutlu konumun kestirilmesi (yöngüdüm (*navigation*))



Erim ölçümleriyle yöngüdüm

- erim ölçümleri $y \in \mathbb{R}^4$, ölçüm gürültüsü v

$$\underbrace{x_d = \begin{bmatrix} 5.59 \\ 10.58 \end{bmatrix}}_{\text{doğru konum}} \qquad \underbrace{y = \begin{bmatrix} -11.95 \\ -2.84 \\ -9.81 \\ 2.81 \end{bmatrix}}_{\text{ölçüm}} \qquad \underbrace{y = \begin{bmatrix} k_1^T \\ k_2^T \\ k_3^T \\ k_4^T \end{bmatrix} x + v}_{\text{model}}$$

- k_i: 0'dan parıldak i'ye doğru olan birim vektör
- yöngüdüm problemi: $y \in \mathbb{R}^4$ verilsin. $x \in \mathbb{R}^2$'yi kestirmek (*estimate*) istiyoruz
- en küçük kareler yöntemi ile kestirim $\hat{x}$'i hesaplayabiliriz:

$$\hat{x} = A^\dagger y = \begin{bmatrix} -0.23 & -0.48 & 0.04 & 0.44 \\ -0.47 & -0.02 & -0.51 & -0.18 \end{bmatrix} y = \begin{bmatrix} 4.95 \\ 10.26 \end{bmatrix}$$