Kısıtlı 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 kısıtlı en küçük kareler

Eşitlik kısıtlarıyla en küçük kareler

- doğrusal **kısıtlı en küçük kareler** (*constrained least squares*) problemi (CLS)

$$\begin{array}{ll} \underset{x}{\text{minimize}} & \|Ax - b\|^2 \\ \text{bağlı} & Cx = d \end{array}$$

  şeklinde bir optimizasyon problemidir
- $x \in \mathbb{R}^n$: değerini bulmak/seçmek istediğimiz vektör
- $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $C \in \mathbb{R}^{p \times n}$, ve $d \in \mathbb{R}^p$ problem verileridir (yani, bunları biliyoruz)
- $\|Ax - b\|^2$: amaç fonksiyonu
- $Cx = d$: eşitlik kısıtları
- $Cx = d$ ise x **olanaklı**dır (*feasible*)
- $C\hat{x} = d$ ise ve $\|A\hat{x} - b\|^2 \leq \|Ax - b\|^2$ koşulu $Cx = d$'yi sağlayan her $x \in \mathbb{R}^n$ için sağlanıyorsa (yani, $\hat{x}$, olanaklı x'ler arasından $\|Ax - b\|^2$'nin en küçük değerini almasını sağlayan seçenek ise), $\hat{x}$ CLS'nin bir çözümüdür

Eşitlik kısıtlarıyla en küçük kareler

- CLS doğrusal denklemlerin (burada, $Cx = d$) çözülmesi ile en küçük kareler problemini (burada, minimize $\|Ax - b\|^2$) birleştirir
- CLS,
  $$\underset{x}{\text{minimize}} \quad \|Ax - b\|^2 + \lambda\|Cx - d\|^2$$
  şeklindeki bir **iki-amaçlı** (*bi-objective*) en küçük kareler probleminde ikincil amacın ağırlığı λ'yı sonsuz yaptığımızda oluşacak optimizasyon problemi olarak düşünülebilir

Bölüm 2

En küçük norm problemi

En küçük norm problemi

- kısıtlı en küçük kareler probleminin ($A = I$, $b = 0$ seçildiğinde oluşan) özel halidir
- en küçük norm problemi

$$\begin{array}{rl} \underset{x}{\text{minimize}} & \|x\|^2 \\ \text{bağlı} & Cx = d \end{array}$$

şeklinde bir optimizasyon problemidir (yazıyla: "bir doğrusal denklem takımını sağlayan vektörler arasından en küçük normlu olanı bul")

Örnek: Hareket planlama

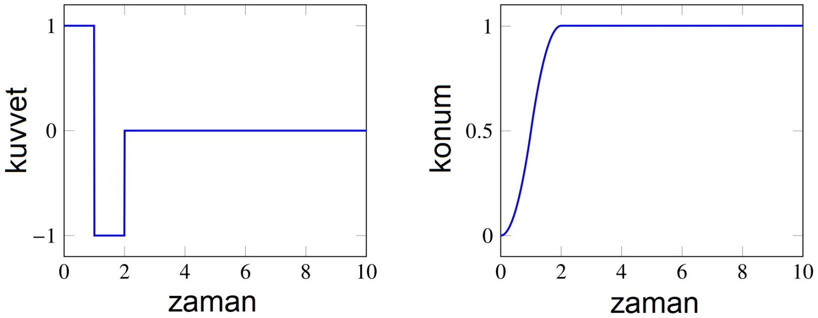
- sürtünmesiz yüzeyde, başlangıçta durgun olan birim kütleli cisim
- model: $F = ma$ (Newton'un ikinci yasası)
- $f \in \mathbb{R}^{10}$: her saniye için cisme uygulanan kuvvet
- cismin nihai (yani, $11.$ saniyedeki) hızı ve konumu:

$$
\begin{aligned}
v^{\text{final}} &= f_1 + f_2 + \cdots + f_{10} \\
p^{\text{final}} &= \frac{19}{2} f_1 + \frac{17}{2} f_2 + \cdots + \frac{1}{2} f_{10}
\end{aligned}
$$

- $v^{\text{final}} = 0$ ve $p^{\text{final}} = 1$ olmasını sağlayacak f vektörünü bulmak istiyoruz
- $f^{\text{bb}} = \begin{bmatrix} 1 & -1 & 0 & \ldots & 0 \end{bmatrix}^T$ bunu sağlar (bu tarz çözümlere **aç-kapa** (*on-off* veya *bang-bang*) denir)

Örnek: Hareket planlama

aç-kapa kuvvet **yörüngesi** (*trajectory*) ve bunun etkisinde oluşan konum yörüngesi



Örnek: Hareket planlama

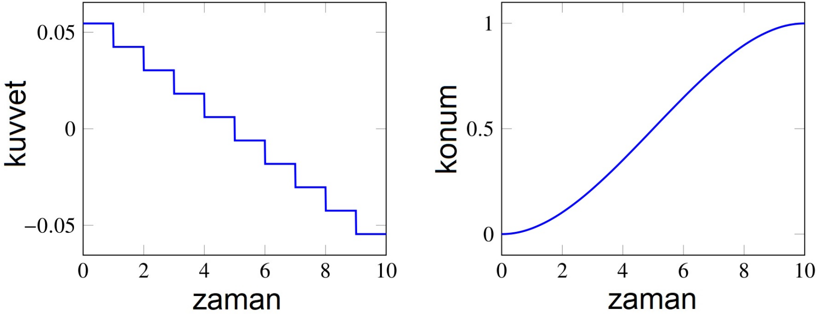
- $v^{\text{final}} = 0$ ve $p^{\text{final}} = 1$ olmasını sağlayacak en küçük normlu f vektörünü bulalım
- en küçük normlu f, pratikte, istenen işi yapan (yani, $v^{\text{final}} = 0$ ve $p^{\text{final}} = 1$ olmasını sağlayan) ve minimum enerji harcaması gerektiren çözüme karşılık gelebilir
- en küçük norm problemi

$$\begin{array}{ll} \underset{f}{\text{minimize}} & \|f\|^2 \\ \text{bağlı} & \begin{bmatrix} 1 & 1 & \cdots & 1 & 1 \\ \frac{19}{2} & \frac{17}{2} & \cdots & \frac{3}{2} & \frac{1}{2} \end{bmatrix} f = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \end{array}$$

- çözüm f^{ln} için amaç fonksiyonunun değeri $\|f^{\text{ln}}\|^2 = 0.0121$ olur (karşılaştırın: $\|f^{\text{bb}}\|^2 = 2$)

Örnek: Hareket planlama

en küçük normlu kuvvet yörüngesi ve bunun etkisinde oluşan konum yörüngesi



Bölüm 3

Kısıtlı en küçük kareler problemini çözmek

Diferansiyel hesapla optimalite koşulları

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f(x) = \|Ax - b\|^2 \\ \text{bağlı} & c_i^T x = d_i, \quad i = 1, \ldots, p \end{array}$$

şeklinde verilen kısıtlı optimizasyon problemini çözmek için:

1. Lagrange fonksiyonunu oluştur (Lagrange çarpanları $z_1, \ldots, z_p$ ile)

$$L(x, z) = f(x) + z_1(c_1^T x - d_1) + \cdots + z_p(c_p^T x - d_p)$$

2. optimalite koşullarını oluştur

$$\begin{aligned} \frac{\partial L}{\partial x_i}(\hat{x}, z) &= 0, \quad i = 1, \ldots, n \\ \frac{\partial L}{\partial z_i}(\hat{x}, z) &= 0, \quad i = 1, \ldots, p \end{aligned}$$

Diferansiyel hesapla optimalite koşulları

- $\frac{\partial L}{\partial z_i}(\hat{x}, z) = c_i^T \hat{x}_i - d_i = 0$ (eşitlik kısıtları için olanaklılık)
- **durağanlık** (*stationarity*)

$$\frac{\partial L}{\partial x_i}(\hat{x}, z) = 2\sum_{j=1}^{n}(A^T A)_{ij}\hat{x}_j - 2(A^T b)_i + \sum_{j=1}^{p} z_j c_i = 0$$

- matris-vektör formunda: $2(A^T A)\hat{x} - 2A^T b + C^T z = 0$
- bunu $C\hat{x} = d$ ile birleştirerek Karush-Kuhn-Tucker (KKT) koşullarını oluşturabiliriz:

$$\underbrace{\begin{bmatrix} 2A^T A & C^T \\ C & 0 \end{bmatrix}}_{\text{KKT matrisi}} \begin{bmatrix} \hat{x} \\ z \end{bmatrix} = \begin{bmatrix} 2A^T b \\ d \end{bmatrix}$$

  bunlar, değişkenleri $\hat{x}$ ve z olan $n + p$ adet denklemden oluşan bir kare denklem takımıdır

- bu denklem takımına "KKT sistemi" de denir

CLS probleminin çözümü

- KKT matrisi tersi alınabilir ise, KKT sisteminin çözümü

$$\begin{bmatrix} \hat{x} \\ z \end{bmatrix} = \begin{bmatrix} 2A^TA & C^T \\ C & 0 \end{bmatrix}^{-1} \begin{bmatrix} 2A^Tb \\ d \end{bmatrix}$$

  şeklinde bulunur
- ancak ve ancak C'nin satırları doğrusal bağımsız ve $\begin{bmatrix} A & C \end{bmatrix}^T$'nin sütunları doğrusal bağımsız ise KKT matrisi tersi alınabilirdir
- bu, $m + p \geq n$ ve $p \leq n$ koşullarını gerektirir
- $\hat{x}$'i hesaplamanın maliyeti: $2mn^2 + 2(n+p)^3$ flop (büyük n için yaklaşık maliyet: n^3)

Çözümün doğrudan teyit edilmesi

- $\hat{x}$'in çözüm olduğunu göstermek için, $Cx = d$'yi sağlayan bir x'i ele alalım
- bu durumda:

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

- son terimi $\big(2A^T(A\hat{x} - b) = -C^T z$ ve $Cx = C\hat{x} = d$ eşitliklerini kullanarak$\big)$ açarsak:

$$\begin{aligned}2(Ax - A\hat{x})^T(A\hat{x} - b) &= 2(x - \hat{x})^T A^T (A\hat{x} - b) \\ &= -(x - \hat{x})^T C^T z \\ &= -(C(x - \hat{x}))^T z = 0\end{aligned}$$

- $\|Ax - b\|^2 = \|A(x - \hat{x})\|^2 + \|A\hat{x} - b\|^2 \geq \|A\hat{x} - b\|^2$
- buradan $\hat{x}$'in çözüm olduğu sonucuna varırız

En küçük norm probleminin çözümü

- en küçük norm problemi:

$$\begin{array}{rl} \underset{x}{\text{minimize}} & \|x\|^2 \\ \text{bağlı} & Cx = d \end{array}$$

- $\begin{bmatrix} I & C \end{bmatrix}^T$'nin sütunları daima doğrusal bağımsızdır
- C'nin satırlarının doğrusal bağımsız olduğunu varsayıyoruz
- optimalite koşulu (KKT sistemi):

$$\begin{bmatrix} 2I & C^T \\ C & 0 \end{bmatrix} \begin{bmatrix} \hat{x} \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ d \end{bmatrix}$$

- burada birinci denklemden $\hat{x} = -(\frac{1}{2})C^T z$ olduğu görülür. bu halde ikinci denklem $-(\frac{1}{2})CC^T z = d$ şeklinde oluşur
- $z = -2(CC^T)^{-1}$'i birinci denklemde yerine yazarsak, çözümü

$$\hat{x} = C^T(CC^T)^{-1}d = C^\dagger d$$

şeklinde elde ederiz (burada $C^\dagger$, C'nin sözde tersi)

En küçük norm probleminin çözümü

C'nin satırları doğrusal bağımsız ise:

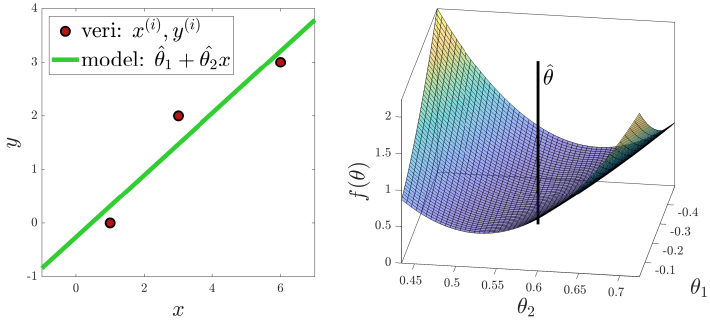
- $C^\dagger$, C'nin bir sağ tersidir
- dolayısıyla her d için $\hat{x} = C^\dagger d$ $C\hat{x} = d$'yi sağlar
- (buradaki analizle şunu öğrendik:) $\hat{x}$, (C geniş matris olduğu için sonsuz adet çözümü olan) $Cx = d$ denkleminin en küçük (normlu) çözümüdür

Bölüm 4

Problemlerin geometrisi

En küçük kareler problemi

örnek: 2 boyutlu veriye düz çizgi uydurma

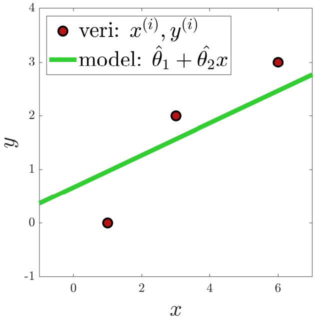


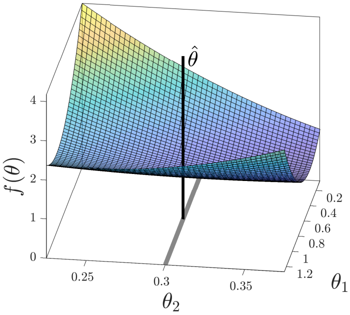
$$A = \begin{bmatrix} 1 & 1 \\ 1 & 3 \\ 1 & 6 \end{bmatrix} \quad b = \begin{bmatrix} 0 \\ 2 \\ 3 \end{bmatrix} \quad \theta = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} \quad f(\theta) = \|A\theta - b\|^2$$

problem: $\underset{\theta}{\text{minimize}}\ f(\theta)$ çözüm: $\hat{\theta} = \begin{bmatrix} -0.263 \\ 0.579 \end{bmatrix}$

Kısıtlı en küçük kareler problemi

örnek: 2 boyutlu veriye düz çizgi uydurma (kısıtlı)





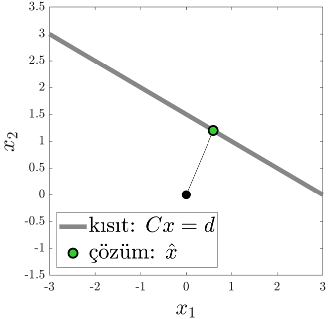
$$A = \begin{bmatrix} 1 & 1 \\ 1 & 3 \\ 1 & 6 \end{bmatrix} \quad b = \begin{bmatrix} 0 \\ 2 \\ 3 \end{bmatrix} \quad \theta = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} \quad f(\theta) = \|A\theta - b\|^2$$

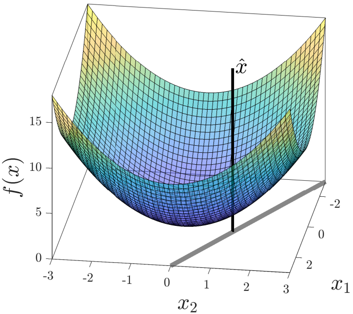
problem: $\underset{\theta}{\text{minimize}} \quad f(\theta)$ çözüm: $\hat{\theta} = \begin{bmatrix} 0.667 \\ 0.3 \end{bmatrix}$

bağlı $\theta_2 = 0.3$

En küçük norm problemi

örnek: kısıtı sağlayan en küçük normlu vektörü bulma





$$C = \begin{bmatrix} 1 & 2 \end{bmatrix} \quad d = 3 \quad x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad f(x) = \|x\|^2$$

problem: $\underset{x}{\text{minimize}} \quad f(x)$ çözüm: $\hat{x} = \begin{bmatrix} 0.6 \\ 1.2 \end{bmatrix}$

bağlı $\quad Cx = d$

Bölüm 5

Uygulamalar

Alt Bölüm 1

Portföy optimizasyonu

Portföy paylaştırma ağırlıkları

- belirli bir periyot (bir gün, hafta, ay vb.) için, n farklı varlığa (hisse, tahvil vb.) toplam V kadar para yatıracağız
- kısa pozisyon almak (yani, bir varlığı başlangıçta ödünç alıp hemen satmak, sürenin sonunda da ödünç aldığımız yere geri vermek) mümkün
- **portföy paylaştırma** (*portfolio allocation*) ağırlık vektörü w, portföyde bulunan varlıkların toplam portföy değerine oranını içeriyor
- Vw_v, varlık j'ye yatırılmış para
- $\mathbf{1}^T w = 1$ sağlanır (w_i pozitifse uzun pozisyon, negatifse kısa pozisyon anlamına gelir)
- örnek: $w = \begin{bmatrix} -0.2 & 0 & 1.2 \end{bmatrix}$ şeklindeki bir ağırlık vektörü, varlık 1 için $0.2V$'lik kısa pozisyon aldığımız, varlık 3 için $1.2V$'lik uzun pozisyon (bu varlıktan alıp elimizde tuttuğumuz) aldığımız, varlık 2 ile ilgili ise herhangi bir pozisyon almadığımız anlamına gelir

Kaldıraç, sadece uzun portföy ve nakit

- **kaldıraç** (*leverage*): $L = |w_1| + \cdots + |w_n|$
- (not: kaldıraç = işlem miktarı/yatırım miktarı)
- bütün ağırlıklar negatif olmayan ise $L = 1$ olur (buna **sadece uzun portföy** (*long only portfolio*) denir)
- $w = \mathbf{1}/n$'e üniform portföy denir

- n. varlık genellikle risksiz varlık (örneğin, nakit para veya hazine bonosu) olarak seçilir
- dolayısıyla, $w = \mathrm{e}_n$ (e_n, n. birim vektör) portföyde sadece nakit para var demektir

Periyot için oluşan getiri

- $\tilde{r}_j$: varlık j'nin periyot için olan **getiri**si (*return*)
- $\tilde{r}_j$, varlığın fiyatonda olan oransal değişim
- genellikle yüzde olarak ifade edilir, örneğin $+2.3\%$ veya -1.1%
- portföyün getirisi (V^+ periyot sonundaki portföy değeri):

$$\frac{V^+ - V}{V} = \tilde{r}^T w$$

- portföyü t periyot için tutarsak (ve getiriler $r_1, \ldots, r_t$ olursa) oluşaral portföy değeri:

$$V_{t+1} = V_1(1 + r_1)(1 + r_2) \cdots (1 + r_t)$$

Getiri matrisi

- ağırlık w ile T periyot için portföy tutalım
- (varlık) getiri matrisi tanımlayalım: $R \in \mathbb{R}^{T \times n}$
- burada R_{tj}, varlık j'nin periyot t'deki getirisi
- R'nin t. satırı $\tilde{r}_t^T$ (burada $\tilde{r}_t$ periyot t için varlık getiri vektörü)
- R'nin j. sütunu, varlık j'nin getirilerinin **zaman serisi** (*time series*)
- portföy getirileri vektörü (zaman serisi): $r = Rw$ (T-vektör)
- n. varlık risksiz ise, R'nin son sütunu $\mu^{\mathrm{rs}}\mathbf{1}$ olur (burada μ^{rs} periyot-başına risksiz faiz oranı)

Portföy getirisi ve risk

- r portföy getirilerinin zaman serisi (vektör)
- ortalama getiri: $\mathrm{avg}(r)$
- risk: $\mathrm{std}(r)$
- bunlar periyor-başına getiri ve risk
- küçük periyot-başına getiriler için

$$\begin{aligned} V_{T+1} &= V_1(1+r_1)\cdots(1+r_2) \\ &\approx V_1 + V_1(r_1+\cdots+r_T) \\ &= V_1 + T\mathrm{avg}(r)V_1 \end{aligned}$$

- dolayısıyla, getiri portföy değerindeki ortalama periyot-başına artışın yaklaşık değerini ifade eder

Yıllık getiri ve risk

- ortalama getiri ve risk genellikle yıllık şekilde (yani, yıl-başına) ifade edilir
- her yıl için P adet **işlem** (*trading*) periyodu varsa:

$$\text{yıllık getiri} = P\,\mathrm{avg}(r), \qquad \text{yıllık risk} = \sqrt{P}\,\mathrm{std}(r)$$

  olur (riskin yıllık ifadesindeki karekök, getirinin ortalaması etrafındaki dalgalanmaların bağımsız olduğu varsayımından kaynaklanır)
- getiriler günlük olsun ve bir yılda 250 işlem günü olsun. bu durumda yıllık getiri ve risk

$$\text{yıllık getiri} = 250\,\mathrm{avg}(r), \qquad \text{yıllık risk} = \sqrt{250}\,\mathrm{std}(r)$$

  şeklinde yazılır

Portföy optimizasyonu

- portföy optimizasyonu problemi: portföy ağırlık vektörü w'yi nasıl seçmeliyiz?
- portföyün (ortalama) getirisi yüksek olsun, riski düşük olsun isteriz

- geçmişte gerçekleşen varlık getirilerini biliyoruz, ancak gelecekte ne olacağını bilmiyoruz
- w'yi, geçmişteki getiriler için iyi başarım gösterecek şekilde seçeceğiz
- beklentimiz, bu şekilde seçtiğimiz w'nin gelecekte de iyi başarım göstermesi (tıpkı veri uydurmada yaptığımız gibi)

Portföy optimizasyonu

$$\begin{array}{ll} \underset{w}{\text{minimize}} & \text{std}(Rw)^2 = \frac{1}{T}\|Rw - \rho\mathbf{1}\|^2 \\ \text{bağlı} & \mathbf{1}^T w = 1 \\ & \text{avg}(Rw) = \rho \end{array}$$

- w: seçmek istediğimiz portföy ağırlık vektörü
- R: geçmişteki varlık getirilerinin getiri matrisi
- Rw: geçmişteki portföy getirileri zaman serileri
- ρ: geçmişteki ortalama getiri (kısıt)
- bu problem formülasyonuyla, belirli (sabit) bir getiri oluşturacak ve riski minimize edecek w'yi seçiyoruz
- çözüm w Pareto optimaldir (yani, çözümden hem getirisi daha yüksek hem de riski daha düşük başka bir vektör bulmak imkansızdır)
- bu problemle aslında "gelecekteki getirileri bilseydik, bunlar için geçerli olacak en iyi sabit paylaştırma hangisi olurdu?" sorusunu soruyoruz

CLS ile portföy optimizasyonu

$$\begin{array}{ll} \underset{w}{\text{minimize}} & \|Rw - \rho\mathbf{1}\|^2 \\ \text{bağlı} & \begin{bmatrix} \mathbf{1}^T \\ \mu^T \end{bmatrix} w = \begin{bmatrix} 1 \\ \rho \end{bmatrix} \end{array}$$

- $\mu = R^T\mathbf{1}/T$ (n-vektör): geçmişteki varlık getirileri
- ρ: geçmişteki ortalama getiri (kısıt)
- bu bir kısıtlı en küçük kareler (CLS) problemidir ve çözümü

$$\begin{bmatrix} w \\ z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} 2R^TR & \mathbf{1} & \mu \\ \mathbf{1}^T & 0 & 0 \\ \mu^T & 0 & 0 \end{bmatrix}^{-1} \begin{bmatrix} 2\rho T\mu \\ 1 \\ \rho \end{bmatrix}$$

şeklinde bulunur

Optimal portföyler

- tek varlığa kıyasla çok daha yüksek başarım gösterirler
- risk-getiri eğrisi bir düz çizgidir
- çizginin bir ucunda risksiz varlık bulunur
- **iki-fon** (*two-fund*) teoremi: optimal portföy w, ρ'nun afin bir fonksiyonudur:

$$\begin{bmatrix} w \\ z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} 2R^T R & \mathbf{1} & \mu \\ \mathbf{1}^T & 0 & 0 \\ \mu^T & 0 & 0 \end{bmatrix}^{-1} \left(\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + \rho \begin{bmatrix} 2T\mu \\ 0 \\ 1 \end{bmatrix} \right)$$

Büyük varsayım

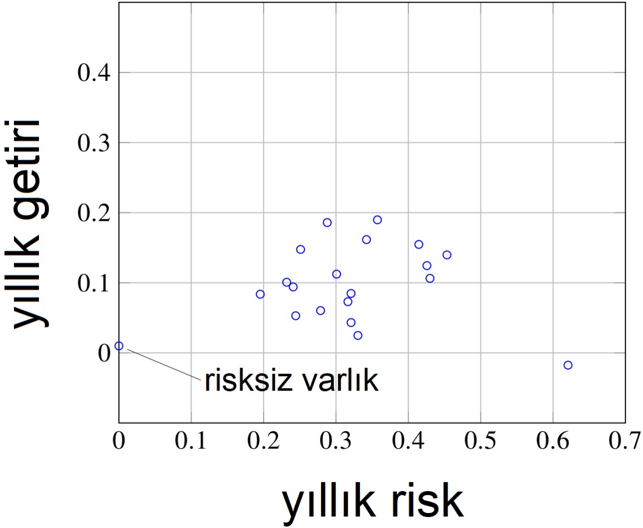
- şimdi bir **büyük varsayım** (*big assumption*) yapıyoruz:

  GELECEK GETİRİLER GEÇMİŞ GETİRİLERE BENZEYECEK

  - her yatırım işleminde, bu varsayımın yanlış olduğu uyarısı yapılır
  - bu varsayım genellikle makul ölçüde doğrudur
  - market **kayma**sı (*shift*) dönemlerinde, varsayım çok daha az doğrudur
- varsayım (yaklaşık olarak da olsa) geçerliyse, geçmişteki getiriler için iyi olan bir ağırlık vektörü w gelecekteki (bilinmeyen) getiriler için de iyi olmalıdır
- örnek:
  - w'yi son 2 yılın getirilerin dayanarak hesapla
  - sonraki (gelecek) 6 ay için bu w'yi kullan

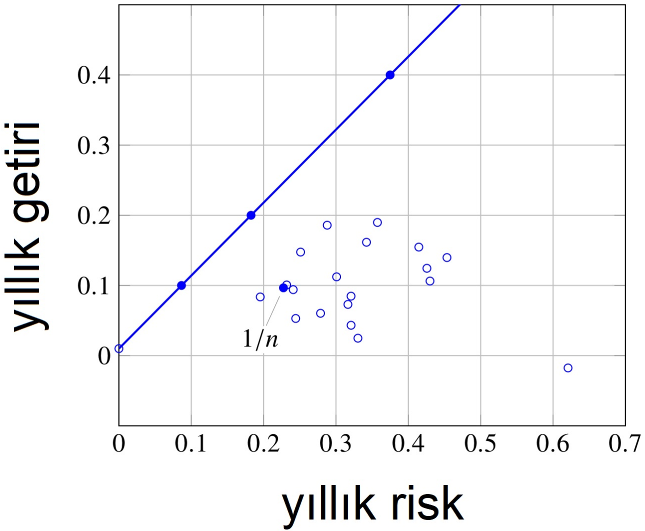
Portföy optimizasyonu - Örnek

veri: 2000 gün boyunca 20 varlık



Portföy optimizasyonu - Örnek

çözüm: pareto optimal portföyler



Portföy optimizasyonu - Örnek

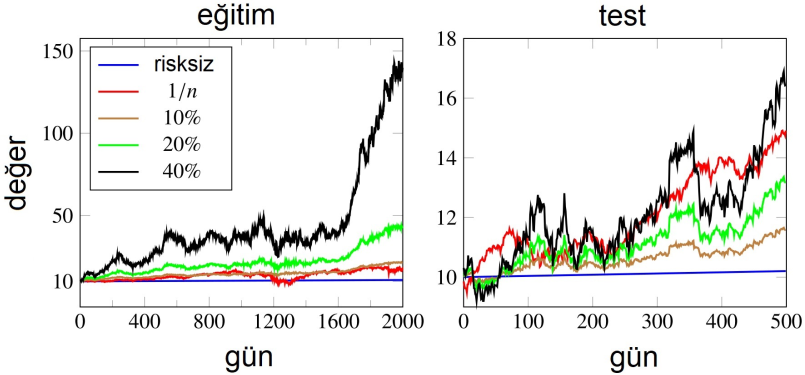
farklı tarzda 5 adet portföyün geçerleme analizi

	getiri		risk		
portföy	eğitim	test	eğitim	test	kaldıraç
risksiz	0.01	0.01	0.00	0.00	1.00
$\rho = 10\%$	0.10	0.08	0.09	0.07	1.96
$\rho = 20\%$	0.20	0.15	0.18	0.15	3.03
$\rho = 40\%$	0.40	0.30	0.38	0.31	5.48
$1/n$ (üniform)	0.10	0.21	0.23	0.13	1.00

- 2000 günlük eğitim periyodu verisini optimal portföyleri hesaplarken kullanıyoruz
- (eğitim verisine dahil olmayan) başka bir 500 günlük test periyodu verisini optimal portföyleri sınarken kullanıyoruz

Portföy optimizasyonu - Örnek

farklı tarzda 5 adet portföyün toplam değeri



Alt Bölüm 2

Doğrusal karesel kontrol

Doğrusal dinamik sistem

$$x_{t+1} = A_t x_t + B_t u_t, \qquad y_t = C_t x_t, \qquad t = 1, 2, \ldots$$

- $x_t \in \mathbb{R}^n$: t anındaki **durum** (*state*)
- $u_t \in \mathbb{R}^m$: t anındaki **giriş** (*input*)
- $y_t \in \mathbb{R}^p$: t anındaki **çıkış** (*output*)
- $A_t \in \mathbb{R}^{n \times n}$: t anındaki durum (veya, dinamik) matrisi
- $B_t \in \mathbb{R}^{n \times m}$: t anındaki giriş matrisi
- $C_t \in \mathbb{R}^{p \times n}$: t anındaki çıkış matrisi
- x_t, u_t ve y_t genellikle standart bir çalışma koşulundan sapmaları temsil eder

Doğrusal karesel kontrol

$$\begin{array}{ll} \underset{\{u_t\}_{t=1}^{t=T-1}}{\text{minimize}} & J_{\text{çıkış}} + \rho J_{\text{giriş}} \\ \text{bağlı} & x_{t+1} = A_t x_t + B_t u_t, \quad t = 1, \ldots, T-1 \\ & x_1 = x^{\text{başlangıç}}, \quad x_T = x^{\text{istenen}} \end{array}$$

- değişkenler: durum yörüngesi $x_1, x_2, \ldots, x_T$ ve giriş yörüngesi $u_1, u_2, \ldots, u_{T-1}$
- iki amaç (yörüngelerin karesel fonksiyonları):

$$J_{\text{çıkış}} = \|y_1\|^2 + \cdots + \|y_T\|^2 = \|C_1 x_1\|^2 + \cdots + \|C_T x_T\|^2$$
$$J_{\text{giriş}} = \|u_1\|^2 + \ldots + \|u_{T-1}\|^2$$

- birinci kısıt yörüngelerin doğrusal dinamik denklemlere uymasını sağlar
- ikinci kısıtlar başlangıç ve nihai (istenen) durumları belirtir
- $\rho \in \mathbb{R}$: iki amaç arasında ödünleşme parametresi (pozitif)

Kısıtlı en küçük kareler formülasyonu

$$\begin{array}{ll} \underset{\{u_t\}_{t=1}^{t=T-1}}{\text{minimize}} & \|C_1x_1\|^2 + \cdots + \|C_Tx_T\|^2 + \rho\left(\|u_1\|^2 + \ldots + \|u_{T-1}\|^2\right) \\ \text{bağlı} & x_{t+1} = A_tx_t + B_tu_t, \quad t = 1, \ldots, T-1 \\ & x_1 = x^{\text{başlangıç}}, \quad x_T = x^{\text{istenen}} \end{array}$$

- doğrusal karesel kontrol problemi

$$\begin{array}{ll} \underset{z}{\text{minimize}} & \|\tilde{A}z - \tilde{b}\|^2 \\ \text{bağlı} & \tilde{C}z = \tilde{d} \end{array}$$

  formunda bir kısıtlı en küçük kareler problemi olarak yazılabilir
- bu formülasyondaki değişken vektörü z, $Tn + (T-1)m$ adet değişken içerir:

$$z = \begin{bmatrix} x_1 & x_2 & \ldots & x_T & u_1 & u_2 & \ldots & u_{T-1} \end{bmatrix}^T$$

Kısıtlı en küçük kareler formülasyonu

$$\tilde{A} = \left[\begin{array}{ccc|ccc} C_1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & C_T & 0 & \cdots & 0 \\ \hline 0 & \cdots & 0 & \sqrt{\rho} I & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & \sqrt{\rho} I \end{array}\right], \qquad \tilde{b} = 0$$

$$\tilde{C} = \left[\begin{array}{cccccc|cccc} A_1 & -I & 0 & \cdots & 0 & 0 & B_1 & 0 & \cdots & 0 \\ 0 & A_2 & -I & \cdots & 0 & 0 & 0 & B_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & A_{T-1} & -I & 0 & 0 & \cdots & B_{T-1} \\ \hline I & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & I & 0 & 0 & \cdots & 0 \end{array}\right], \qquad \tilde{d} = \left[\begin{array}{c} 0 \\ 0 \\ \vdots \\ 0 \\ \hline x^{\text{init}} \\ x^{\text{des}} \end{array}\right]$$

Doğrusal karesel kontrol - Örnek

- sistem **zamanla değişmeyen** (*time-invariant*) (yani, dinamik modeldeki A, B ve C matrisleri sabit):

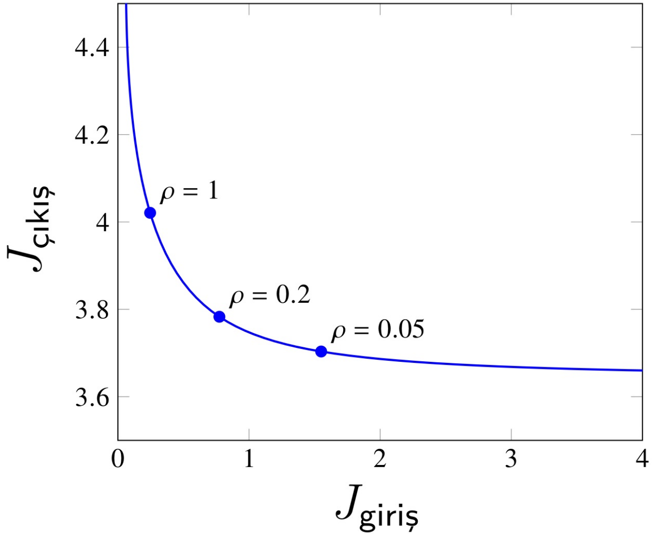
$$A = \begin{bmatrix} 0.855 & 1.161 & 0.667 \\ 0.015 & 1.073 & 0.053 \\ -0.084 & 0.059 & 1.022 \end{bmatrix} \qquad B = \begin{bmatrix} -0.076 \\ -0.0139 \\ 0.342 \end{bmatrix}$$

$$C = \begin{bmatrix} 0.218 & -3.597 & -1.683 \end{bmatrix}$$

- başlangıç koşulu: $x^{\text{başlangıç}} = \begin{bmatrix} 0.496 & -0.745 & 1.394 \end{bmatrix}^T$
- istenen (veya hedef) nihai durum: $x^{\text{istenen}} = 0$
- $T = 100$ (zaman ufku)

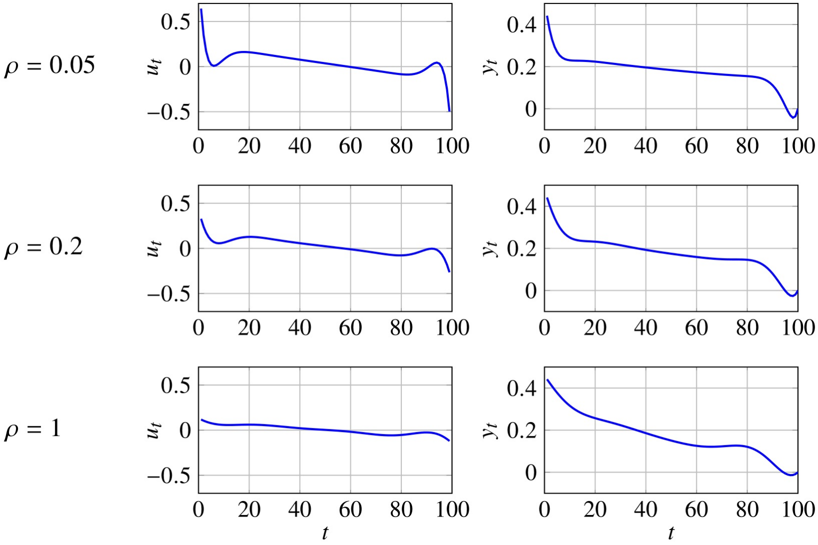
Doğrusal karesel kontrol - Örnek

optimal ödünleşim eğrisi



Doğrusal karesel kontrol - Örnek

optimal ödünleşim eğrisi üzerindeki üç nokta için yörüngeler



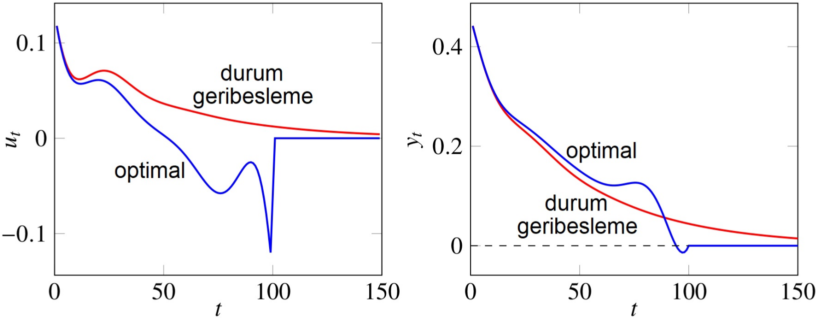
Doğrusal durum geribeslemeli kontrol

- doğrusal durum geribeslemeli kontrolde kontrol girişi
$$u_t = Kx_t, \quad t = 1, 2, \ldots$$
şeklinde hesaplanır
- $K \in \mathbb{R}^{m \times n}$: durum geribesleme kazanç matrisi
- özellikle x_t'nin 0'a yakınsamasının istendiği ve T'nin belirtilmediği hallerde yaygın kullanılan bir yöntemdir

- K'yı seçmek için bir yöntem: $x^{\text{istenen}} = 0$ ile doğrusal karesel kontrol problemini çözmek
- çözüm u_t, x^{istenen}'in bir doğrusal fonksiyonudur, dolayısıyla u_1, $u_1 = Kx^{\text{başlangıç}}$ şeklinde yazılabilir
- K'nın sütunları u_1'yu $x^{\text{başlangıç}} = \mathrm{e}_1, \ldots, \mathrm{e}_n$ için hesaplayarak bulunabilir
- bu şekilde oluşturulan K durum geribesleme kazanç matrisi olarak kullanılabilir

Doğrusal durum geribeslemeli kontrol

örnek: (önceki örnekten devam)



- dinamik modelin matrisleri önceki örnek ile aynı
- mavi eğride optimal doğrusal karesel kontrol ($T = 100$) kullanılıyor
- kırmızı eğride doğrusal durum geribesleme $u_t = Kx_t$ kullanılıyor

Alt Bölüm 3

Doğrusal karesel durum kestirme

Durum kestirme

$$x_{t+1} = A_t x_t + B_t w_t, \qquad y_t = C_t x_t + v_t, \qquad t = 1,\, 2,\, \ldots$$

- $x_t \in \mathbb{R}^n$: t anındaki **durum** (*state*)
- $y_t \in \mathbb{R}^p$: t anındaki **ölçüm** (*measurement*)
- $w_t \in \mathbb{R}^m$: t anındaki **süreç gürültüsü** (*process noise*) (veya, giriş gürültüsü)
- $v_t \in \mathbb{R}^p$: t anındaki **ölçüm gürültüsü** (*measurement noise*) (veya, ölçüm kalıntısı)
- genellikle pratikte w_t **modelleme belirsizliği**ni (*modeling uncertainty*), v_t ise **işaret belirsizliği**ni (*signal uncertainty*) temsil eder
- $A_t \in \mathbb{R}^{n\times n}$, $B_t \in \mathbb{R}^{n\times m}$, $C_t \in \mathbb{R}^{p\times n}$ matrislerini (dinamik modeli) ve ölçümleri $(y_1, y_2, \ldots, y_T)$ biliyoruz
- gürültüleri (w_t ve v_t) bilmiyoruz, ancak bunların küçük olduğunu varsayıyoruz
- **durum kestirme** (*state estimation*) problemi: $x_1, x_2, \ldots, x_T$'yi kestirmek/tahmin etmek

En küçük kareler durum kestirme

$$\begin{array}{ll} \underset{\{w_t\}_{t=1}^{t=T-1}}{\text{minimize}} & J_{\text{ölçüm}} + \lambda J_{\text{süreç}} \\ \text{bağlı} & x_{t+1} = A_t x_t + B_t w_t, \quad t = 1, \ldots, T-1 \end{array}$$

- değişkenler: durum yörüngesi $x_1, x_2, \ldots, x_T$ ve süreç gürültüsü yörüngesi $w_1, w_2, \ldots, w_{T-1}$
- iki amaç (yörüngelerin karesel fonksiyonları):

$$J_{\text{ölçüm}} = \|\underbrace{C_1 x_1 - y_1}_{v_1}\|^2 + \cdots + \|\underbrace{C_T x_T - y_T}_{v_T}\|^2$$

$$J_{\text{süreç}} = \|w_1\|^2 + \ldots + \|w_{T-1}\|^2$$

- $J_{\text{ölçüm}}$: ölçüm kalıntılarının karelerinin toplamı
- $J_{\text{süreç}}$: süreç gürültülerinin karelerinin toplamı
- $\lambda \in \mathbb{R}$: iki amaç arasında ödünleşme parametresi (pozitif)

Kısıtlı en küçük kareler formülasyonu

$$
\begin{array}{ll}
\underset{\{w_t\}_{t=1}^{t=T-1}}{\text{minimize}} & \|C_1x_1 - y_1\|^2 + \cdots + \|C_Tx_T - y_T\|^2 + \ldots \\
& \lambda\left(\|w_1\|^2 + \ldots + \|w_{T-1}\|^2\right) \\
\text{bağlı} & x_{t+1} = A_tx_t + B_tw_t, \quad t = 1, \ldots, T-1
\end{array}
$$

- doğrusal karesel durum kestirme problemi

$$
\begin{array}{ll}
\underset{z}{\text{minimize}} & \|\tilde{A}z - \tilde{b}\|^2 \\
\text{bağlı} & \tilde{C}z = \tilde{d}
\end{array}
$$

  formunda bir kısıtlı en küçük kareler problemi olarak yazılabilir
- bu formülasyondaki değişken vektörü z, $Tn + (T-1)m$ adet değişken içerir:

$$
z = \begin{bmatrix} x_1 & x_2 & \ldots & x_T & w_1 & w_2 & \ldots & w_{T-1} \end{bmatrix}^T
$$

Kısıtlı en küçük kareler formülasyonu

$$\tilde{A} = \left[\begin{array}{cccc|ccc} C_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & C_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & C_T & 0 & \cdots & 0 \\ \hline 0 & 0 & \cdots & 0 & \sqrt{\lambda} I & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & \sqrt{\lambda} I \end{array}\right], \qquad \tilde{b} = \left[\begin{array}{c} y_1 \\ y_2 \\ \vdots \\ y_T \\ \hline 0 \\ \vdots \\ 0 \end{array}\right]$$

$$\tilde{C} = \left[\begin{array}{cccccc|cccc} A_1 & -I & 0 & \cdots & 0 & 0 & B_1 & 0 & \cdots & 0 \\ 0 & A_2 & -I & \cdots & 0 & 0 & 0 & B_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & A_{T-1} & -I & 0 & 0 & \cdots & B_{T-1} \end{array}\right], \qquad \tilde{d} = 0$$

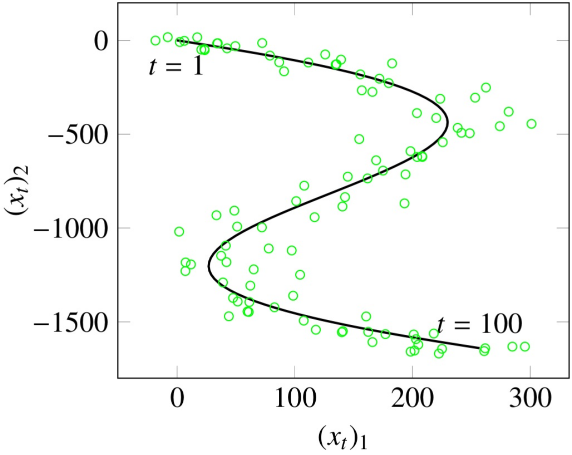
Doğrusal karesel durum kestirme - Örnek

$$A_t = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad B_t = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \quad C_t = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

- 2 boyutlu uzayda hareket eden cismin basit modeli
- $x_t = \begin{bmatrix} p_t & z_t \end{bmatrix}$, $p_t \in \mathbb{R}^2$ konum, $z_t \in \mathbb{R}^2$ hız
- $y_t = C_t x_t + w_t$ ($y_t \in \mathbb{R}^2$): konumun gürültülü ölçümü
- $T = 100$ (zaman ufku)

Doğrusal karesel durum kestirme - Örnek

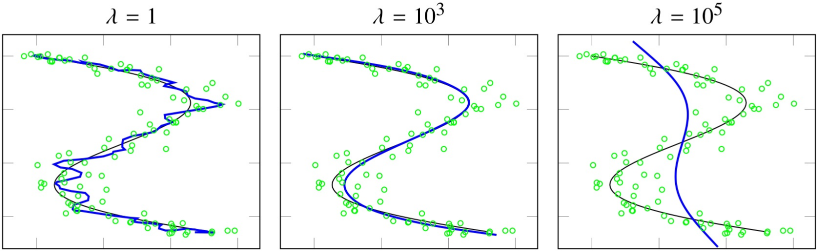
ölçümlerin ve gerçek konumlar



- siyah eğri: cismin gerçek konumu $p_t = C_t x_t$
- yeşil daireler: (100 adet) gürültülü ölçüm $y_t = C_t x_t + v_t$

Doğrusal karesel durum kestirme - Örnek

konum kestirim (*position estimate*) yörüngeleri



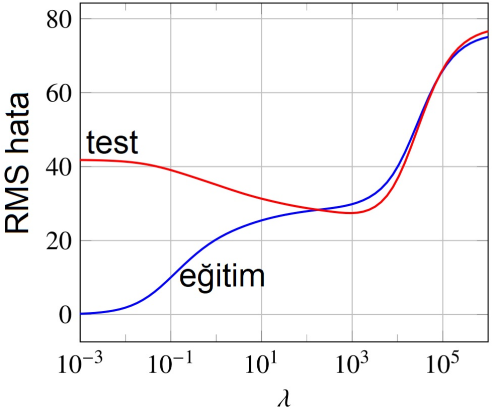
mavi eğriler: üç farklı λ değeri için konum kestirim yörüngeleri
yorum: λ'yı küçük seçmek "veriye (y_t'ye) (modele kıyasla daha fazla) güveniyoruz, dolayısıyla ölçüm gürültüsü v_t'nin (w_t'ye kıyasla) küçük olacağını tahmin ediyoruz", büyük seçmek ise "modele ($x_{t+1} = A_t x_t$'ye) (veriye kıyasla daha fazla) güveniyoruz, dolayısıyla süreç gürültüsü w_t'nin (v_t'ye kıyasla) küçük olacağını tahmin ediyoruz" anlamına geliyor

Çapraz geçerleme

durum kestirme için çapraz geçerleme prosedürü:

- ölçümlerin (örneğin) $\%20$'lik kısmını rastgele çıkaralım ve test veri kümesi olarak, kalan kısmını da eğitim veri kümesi olarak kullanalım
- λ'nın birçok farklı değeri için eğitim veri kümesi için durum kestirme yapalım
- her λ değeri için, test veri kümesi için ölçüm kalıntılarının RMS değerini hesaplayalım
- λ'yı test veri kümesi için kalıntıların RMS değerini minimize edecek şekilde seçelim

Çapraz geçerleme - Örnek



- önceki örneğe çapraz geçerleme yöntemini uyguladık
- 100 ölçümden 20 tanesi test veri kümesi olarak ayrıldı
- grafikten λ'yı yaklaşık olarak 10^3 seçmenin iyi sonuç vereceği anlaşılıyor