

Boğazda Yapay Öğrenme İsmail Arı Yaz Okulu 2018

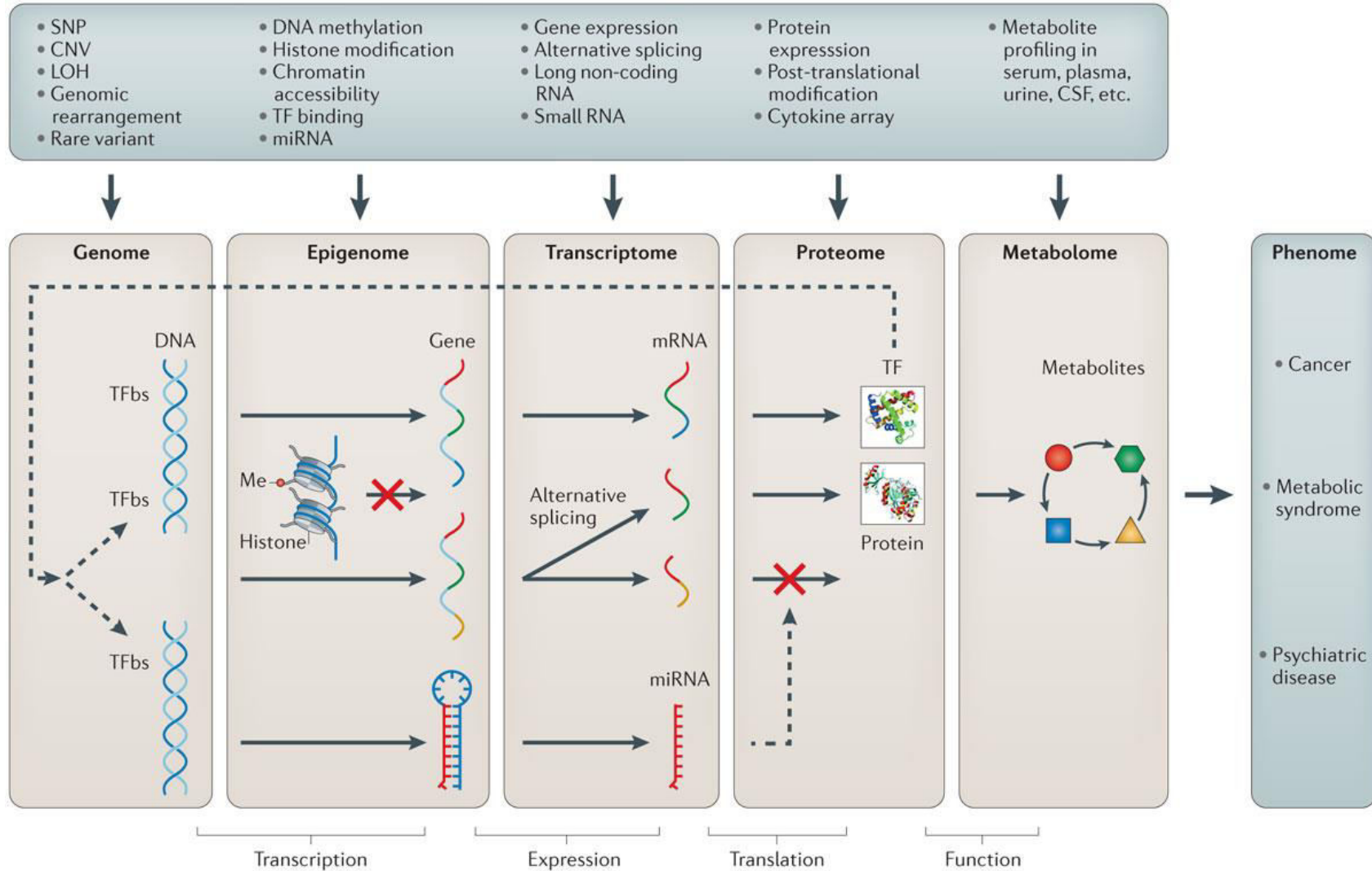
Sıralama Öğrenme ile Sağkalım Tahminleme

Öznur Taştan



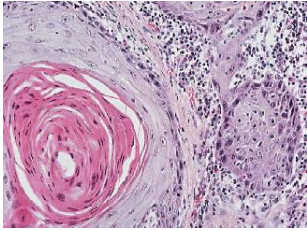
Mühendislik ve Doğa Bilimleri Fakültesi
Bilgisayar Bilimi ve Mühendisliği
Moleküler Biyoloji, Genetik ve Biyomühendislik

Omiks Çağı



Kanser Genom Atlası Projesi

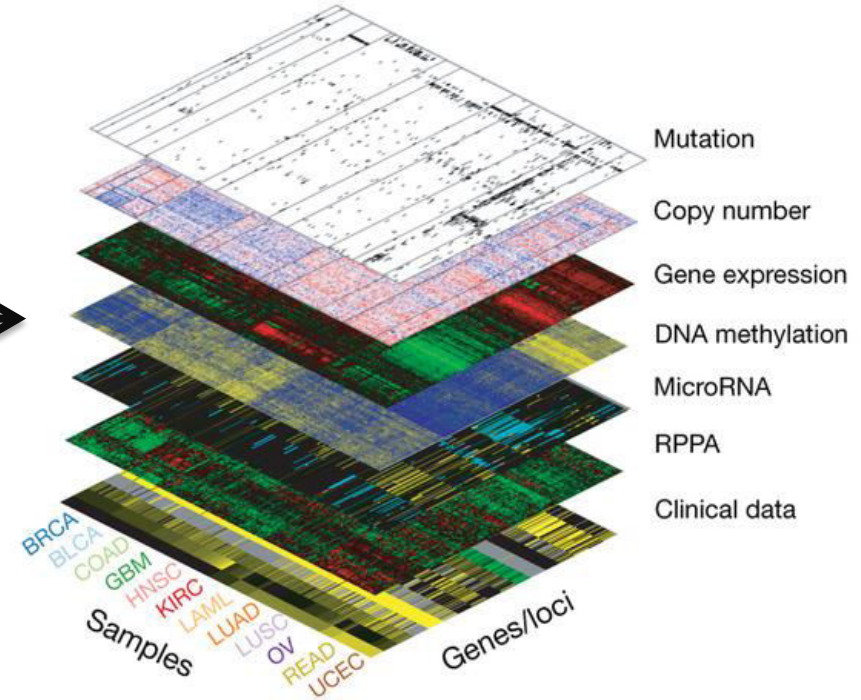
Tümör örnekleri



Deneyler



Omiks Karakterizasyonu

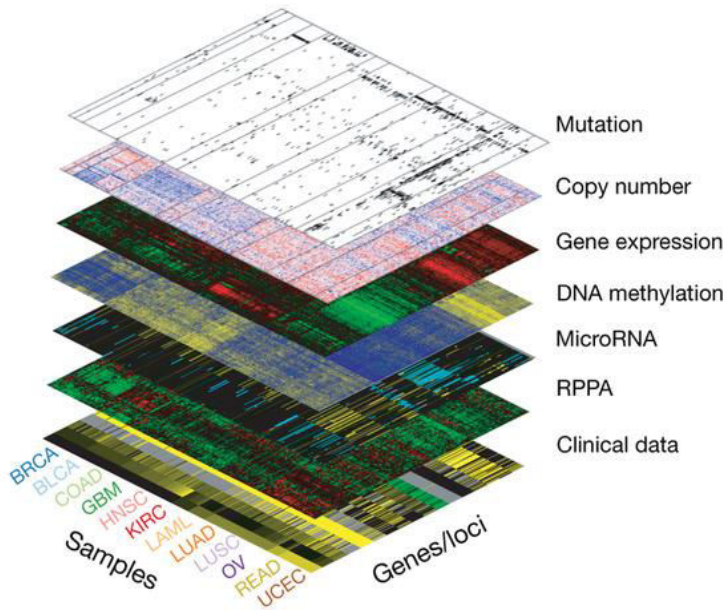


Weinstein et al., Nature Genetics, 2013.

Hasta Sağkalımını Tahminleme

$$x_i \in \mathbf{R}^d$$

Omics characterizations



Weinstein et al., Nature Genetics, 2013.

Sağ kalım süresi ile
omik verileri arasındaki ilişkiyi
özetleyen
bir fonksiyon öğren f



Bu fonksiyonu kullanarak yeni
gelen hastaların omik verisi ile, x_i ,
ne kadar hayatta kalacağını
tahminle

$$T_i^* = f(x_i)$$

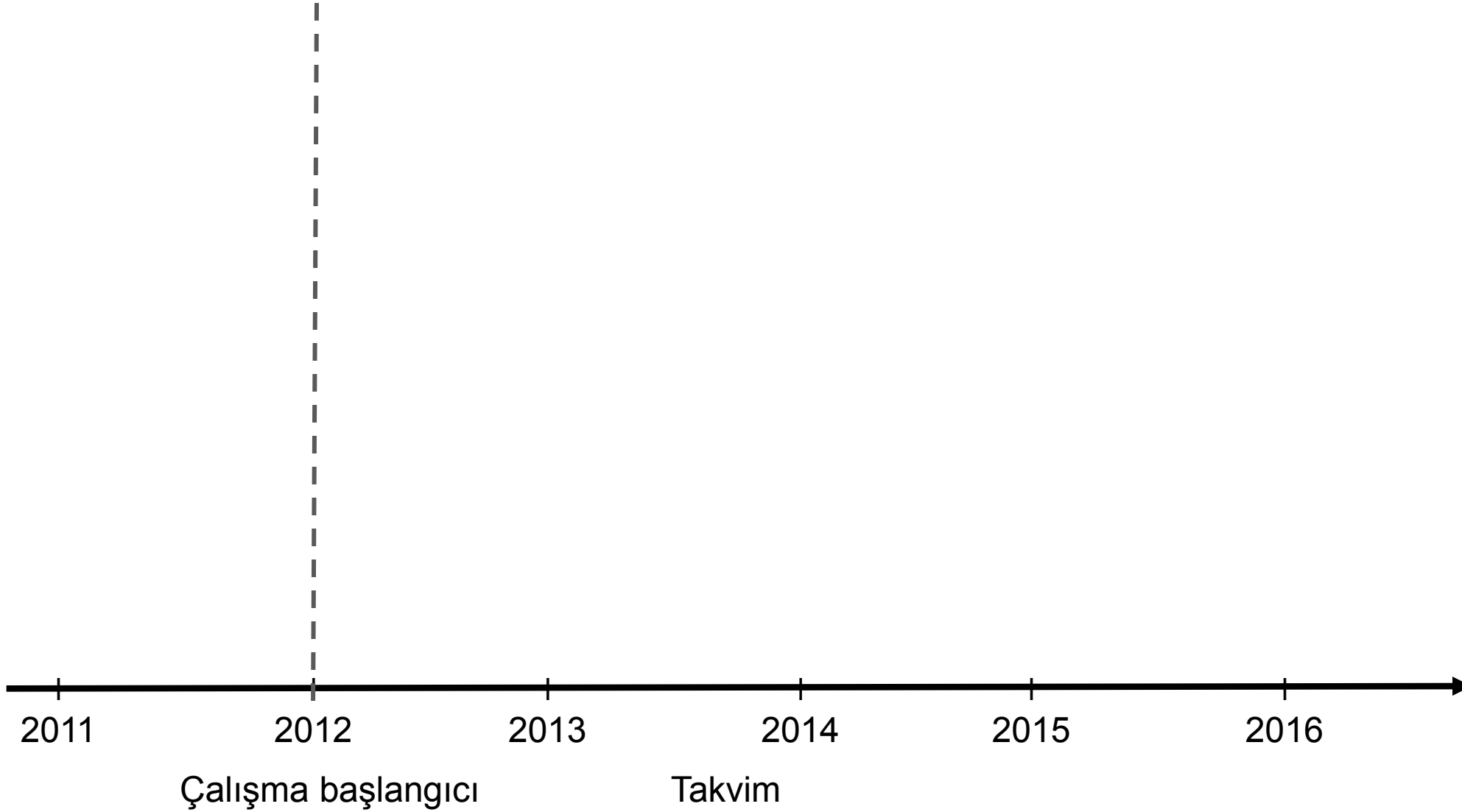
Sağkalım Süresi

- Sağkalım süresi, belirli bir başlangıç zamanından ilgilenilen bir son noktaya kadar geçen süreyi ölçen değişkeni ifade eder.
 - Ölüme kadar geçen süre
 - Hastalığın nüksüne kadar geçen süre
 - Bir makinanın bozulmasına kadar geçen süre

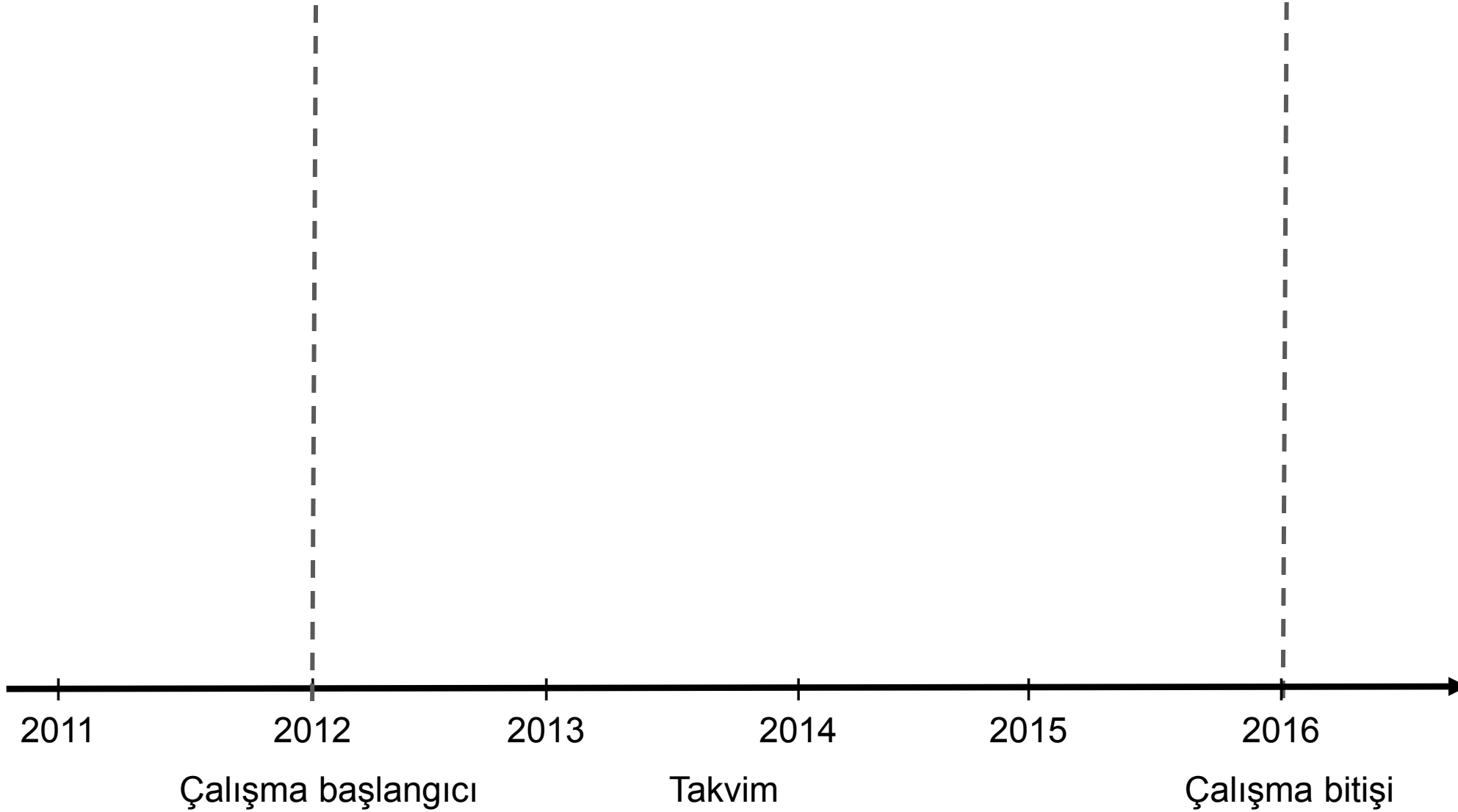
Sağkalım Süresi



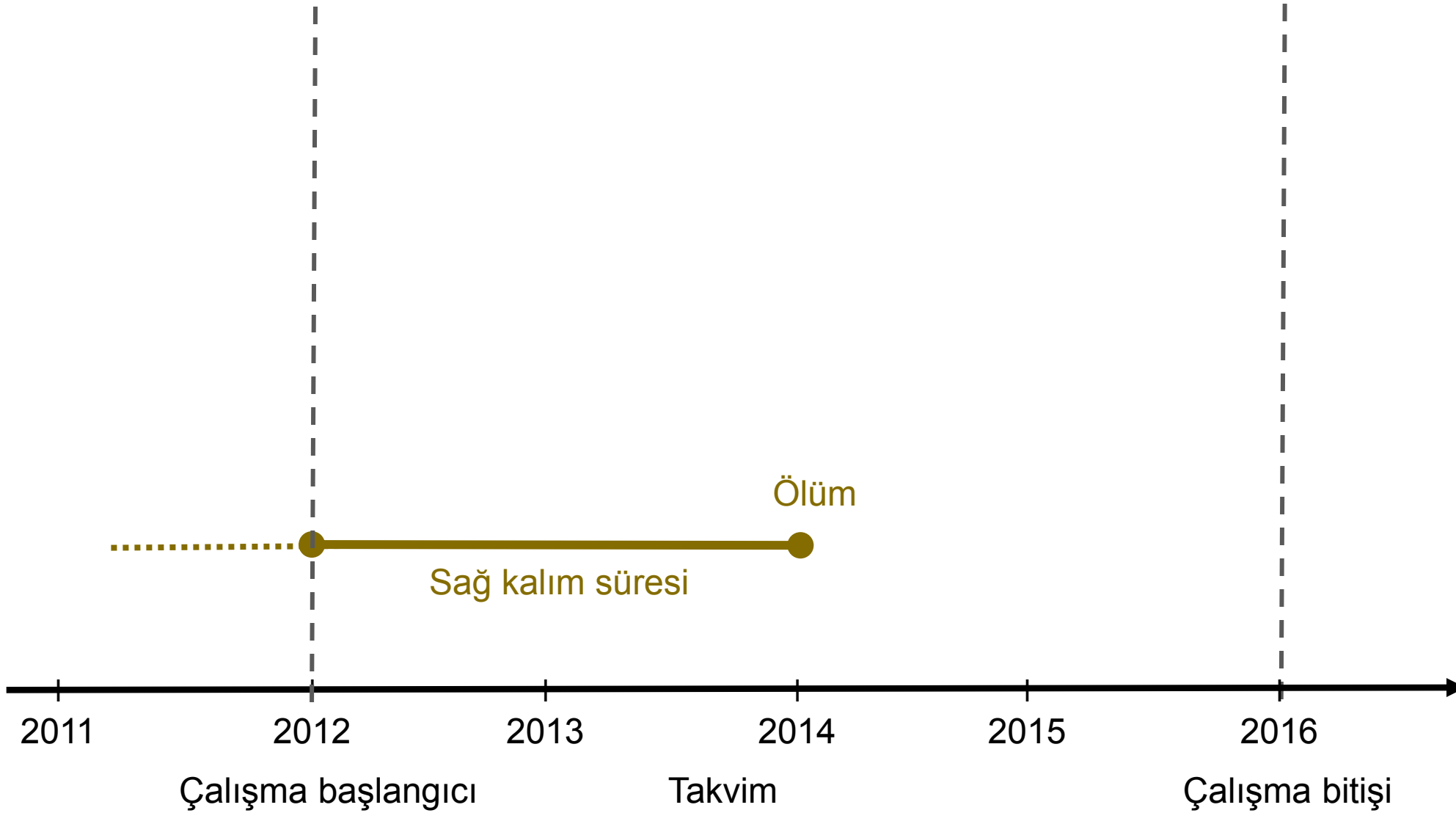
Sağkalım Süresi



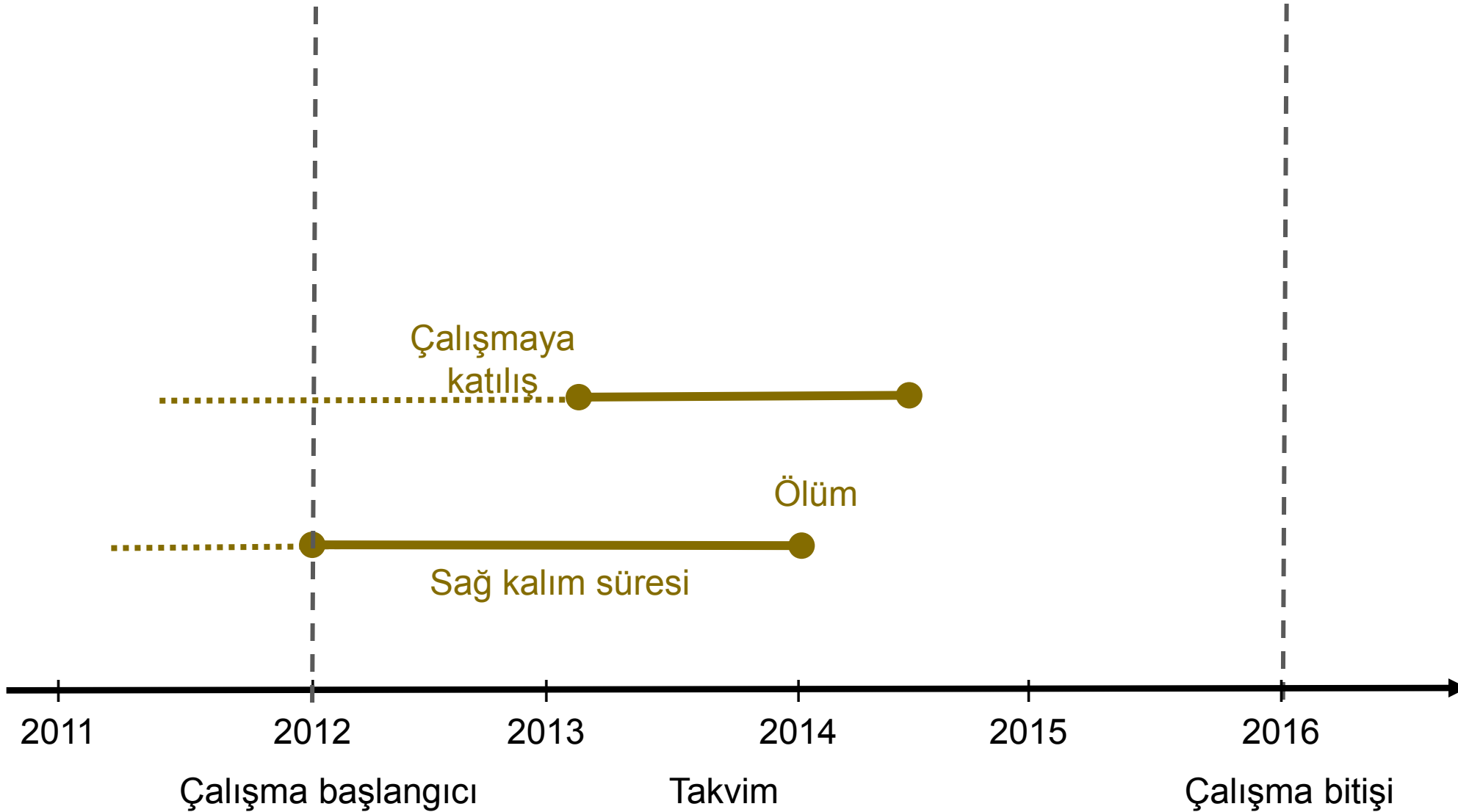
Sağkalım Süresi



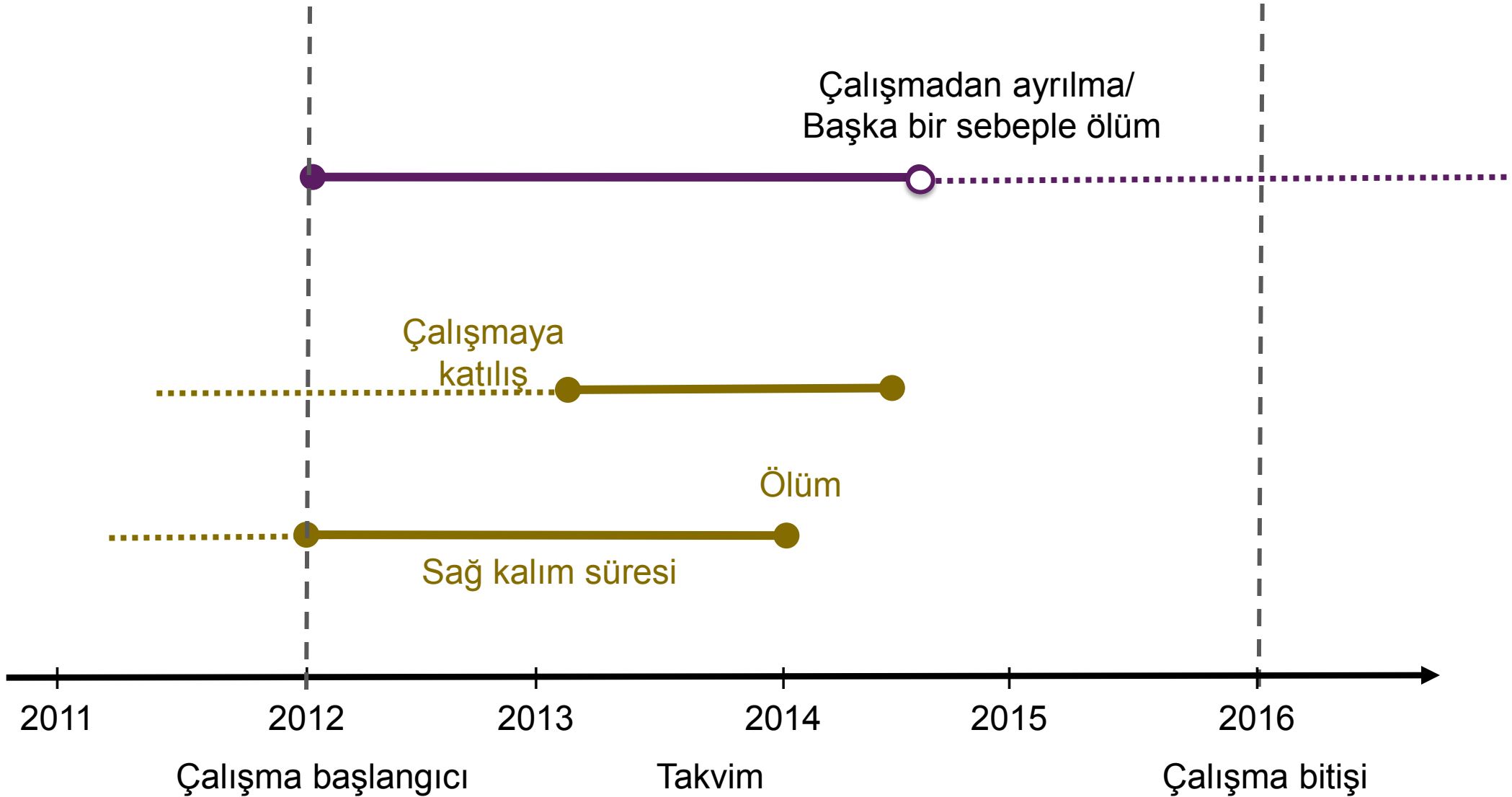
Sağkalım Süresi



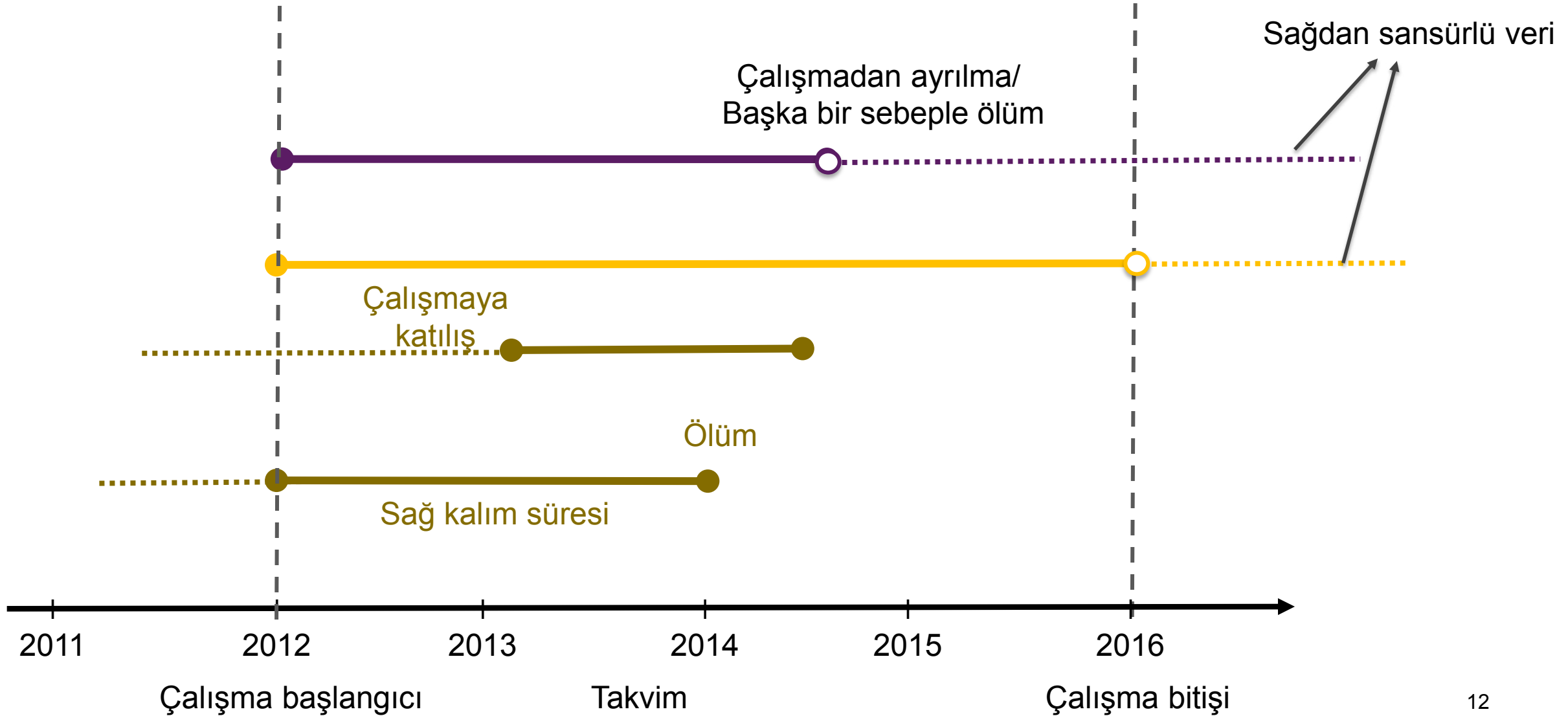
Sağkalım Süresi



Sansürlü Veri



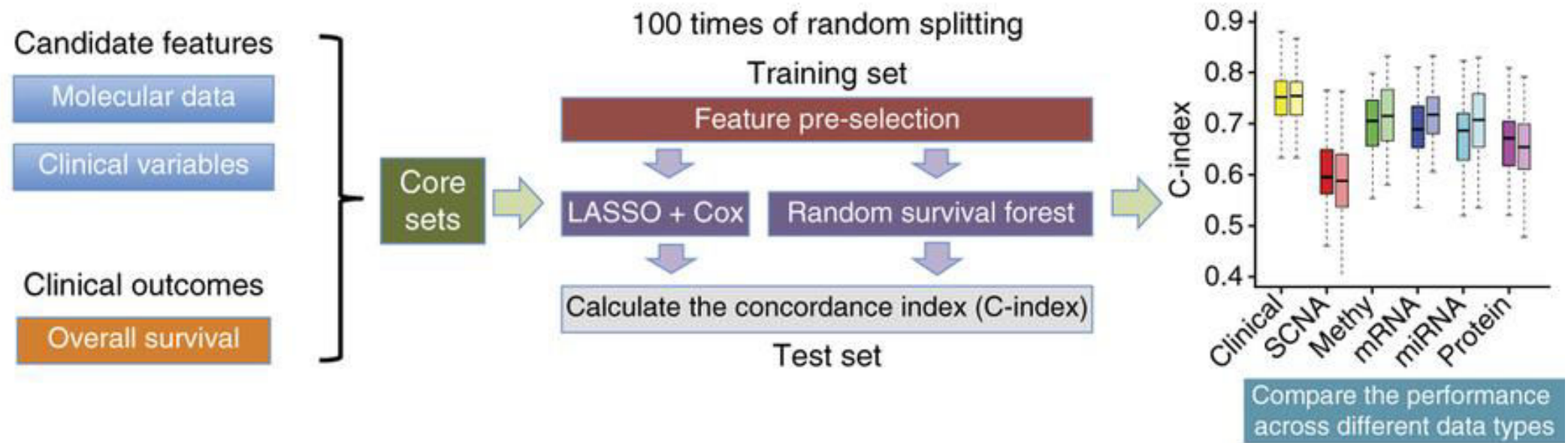
Sansürlü Veri



Assessing the clinical utility of cancer genomic and proteomic data across tumor types

Yuan Yuan^{1,2,14}, Eliezer M Van Allen^{3,4,14}, Larsson Omberg^{5,14}, Nikhil Wagle^{3,4}, Ali Amin-Mansour⁴, Artem Sokolov⁶, Lauren A Byers⁷, Yanxun Xu⁸, Kenneth R Hess⁹, Lixia Diao², Leng Han², Xuelin Huang⁹, Michael S Lawrence⁴, John N Weinstein^{2,10}, Josh M Stuart⁶, Gordon B Mills¹⁰, Levi A Garraway^{3,4,11,15}, Adam A Margolin^{5,13,15}, Gad Getz^{4,11,12,15} & Han Liang^{1,2,15}

Yuan ve ark. Nature Biotechnology 32, 644–652 (2014)



Cox Dr., *J. R. Stat. Soc. Ser. B*, 34, 187-220, 1972.

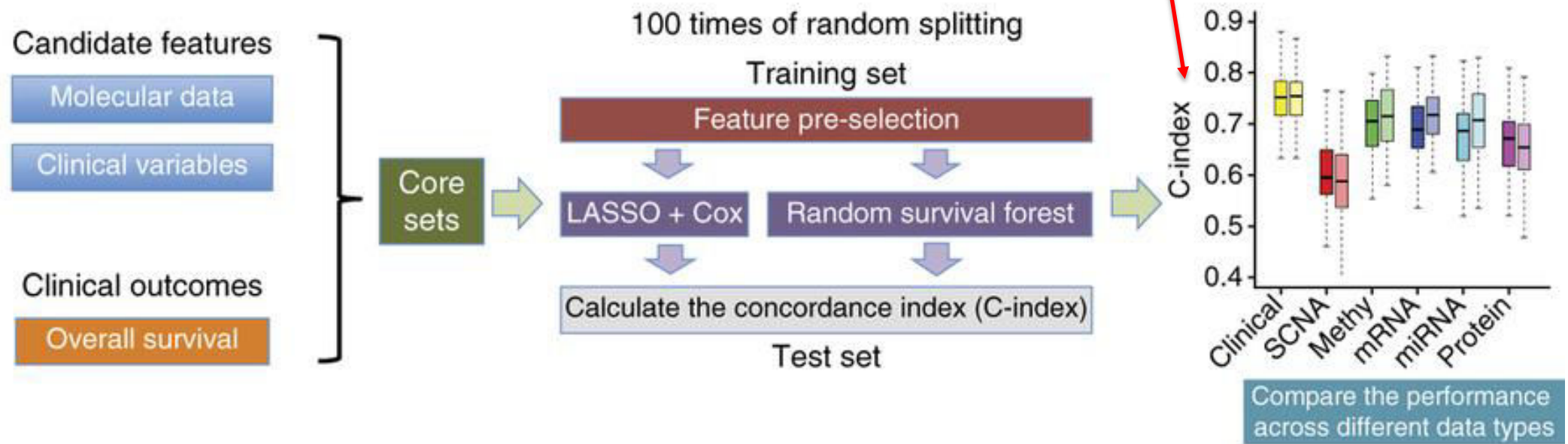
Iswaharan ve ark. *The Annals of Applied Statistics*, 2008.

Assessing the clinical utility of cancer genomic and proteomic data across tumor types

Yuan Yuan^{1,2,14}, Eliezer M Van Allen^{3,4,14}, Larsson Omberg^{5,14}, Nikhil Wagle^{3,4}, Ali Amin-Mansour⁴, Artem Sokolov⁶, Lauren A Byers⁷, Yanxun Xu⁸, Kenneth R Hess⁹, Lixia Diao², Leng Han², Xuelin Huang⁹, Michael S Lawrence⁴, John N Weinstein^{2,10}, Josh M Stuart⁶, Gordon B Mills¹⁰, Levi A Garraway^{3,4,11,15}, Adam A Margolin^{5,13,15}, Gad Getz^{4,11,12,15} & Han Liang^{1,2,15}

Uyum indeksi

Yuan ve ark. Nature Biotechnology 32, 644–652 (2014)



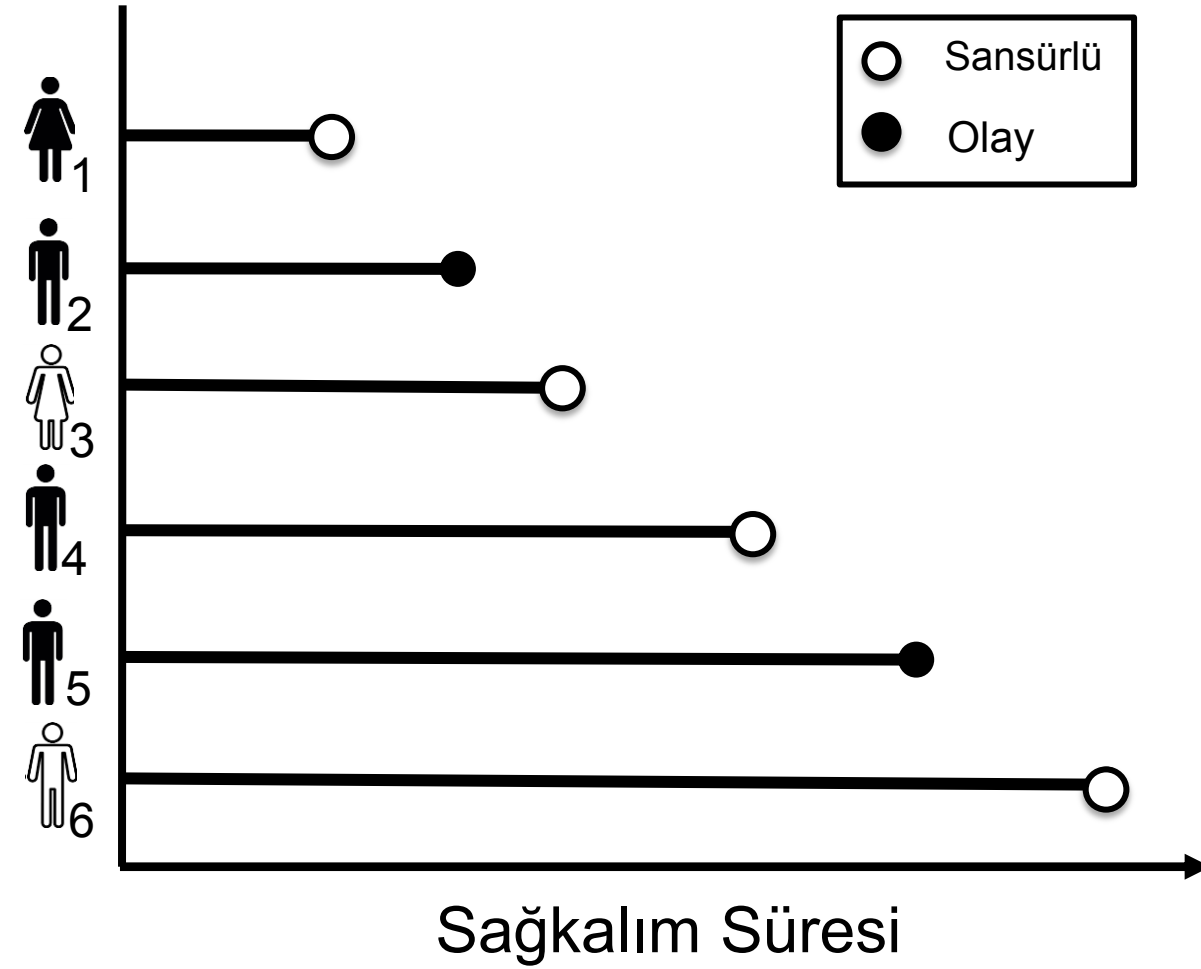
Cox Dr., *J. R. Stat. Soc. Ser. B*, 34, 187-220, 1972.

Iswaharan ve ark. *The Annals of Applied Statistics*, 2008.

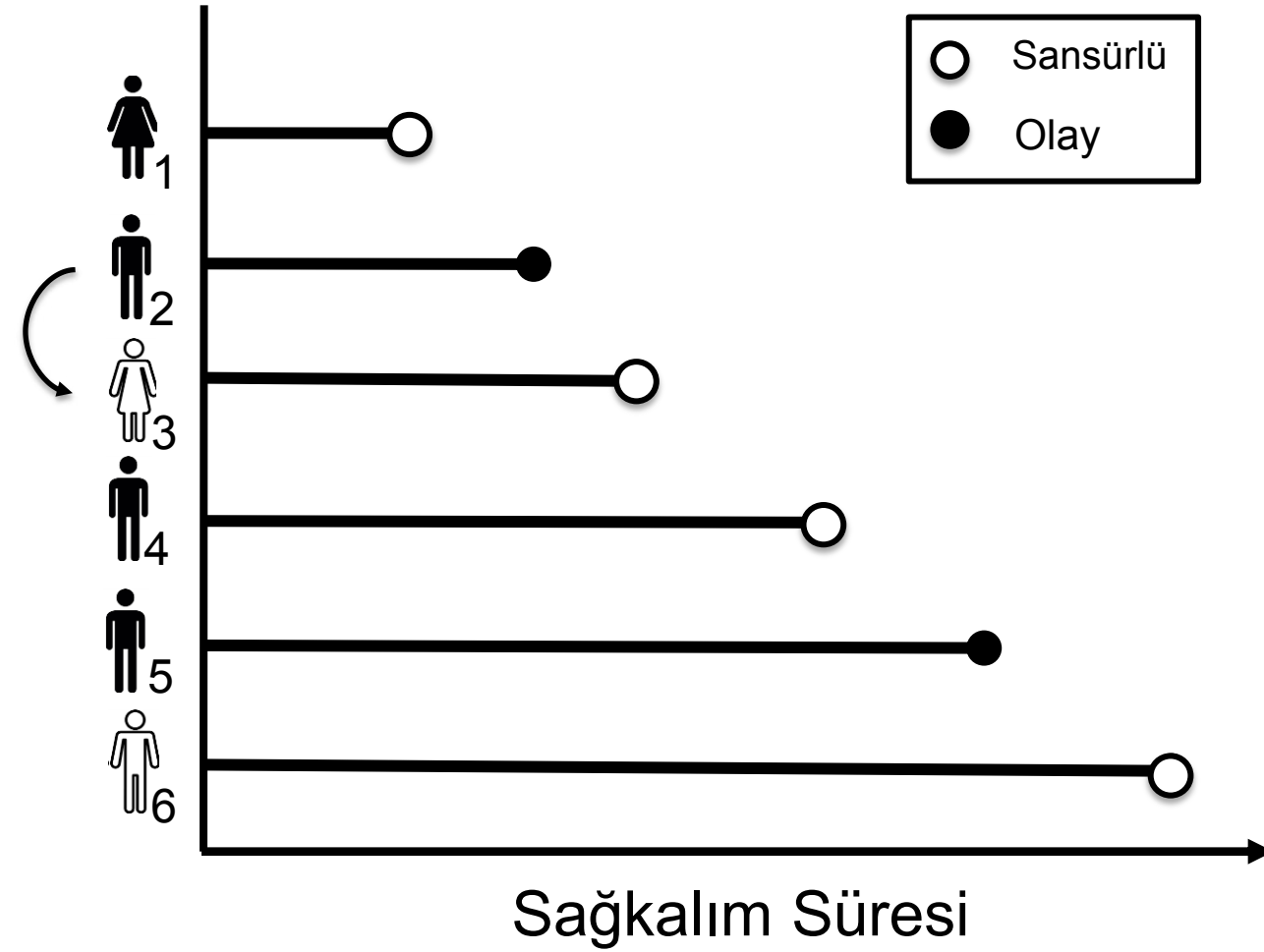
Uyum İndeksi(C-Index)

- Sağkalım tahminlemede kullanılan **standart performans** ölçütü
- Gerçekte var olan hasta sağkalım **sıralaması** ile **tahmin edilen** hasta sağkalım **sıralaması** arasındaki uyum

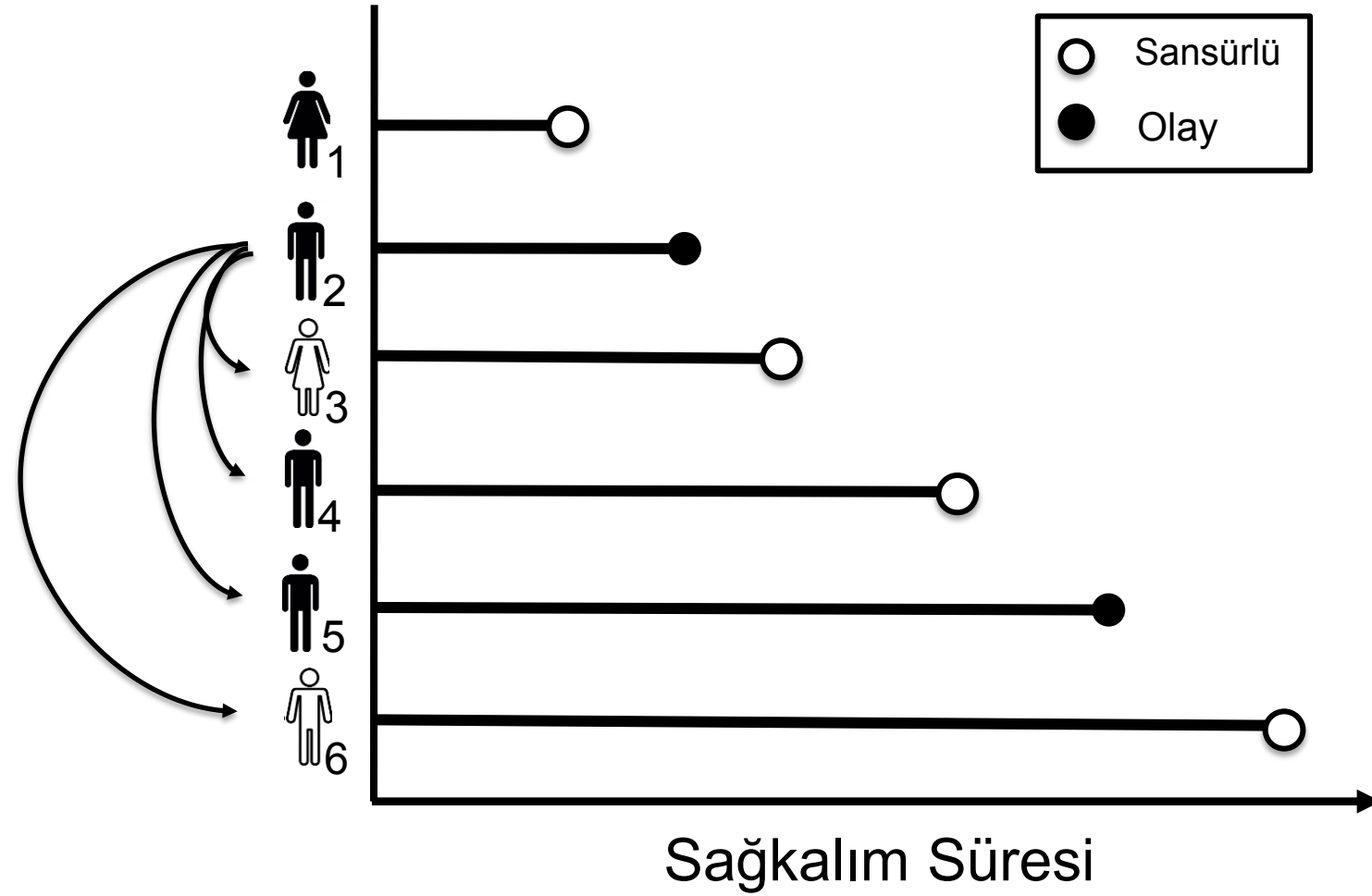
Uyum İndeksi(C-Index)



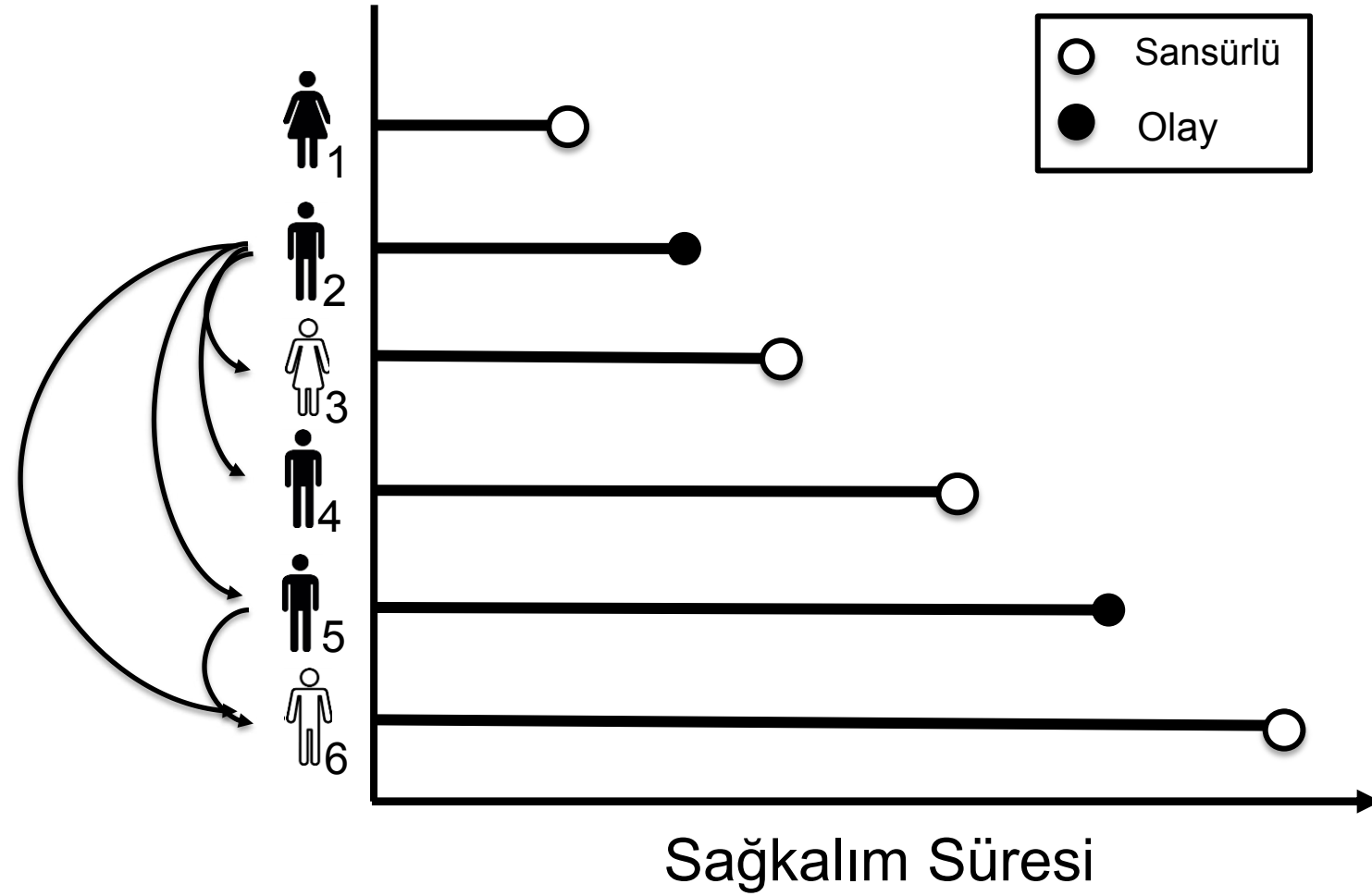
Uyum İndeksi(C-Index)



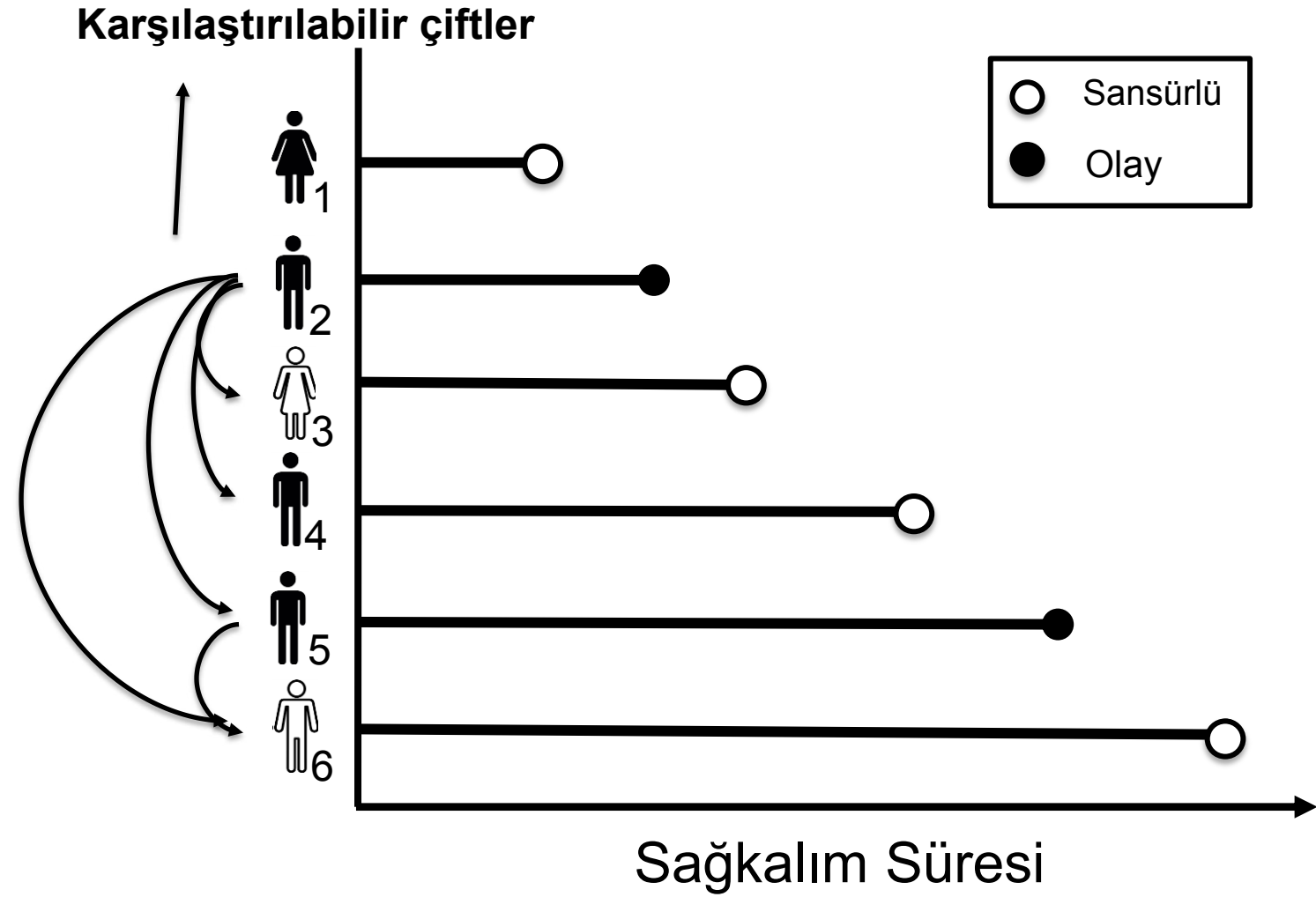
Uyum İndeksi(C-Index)



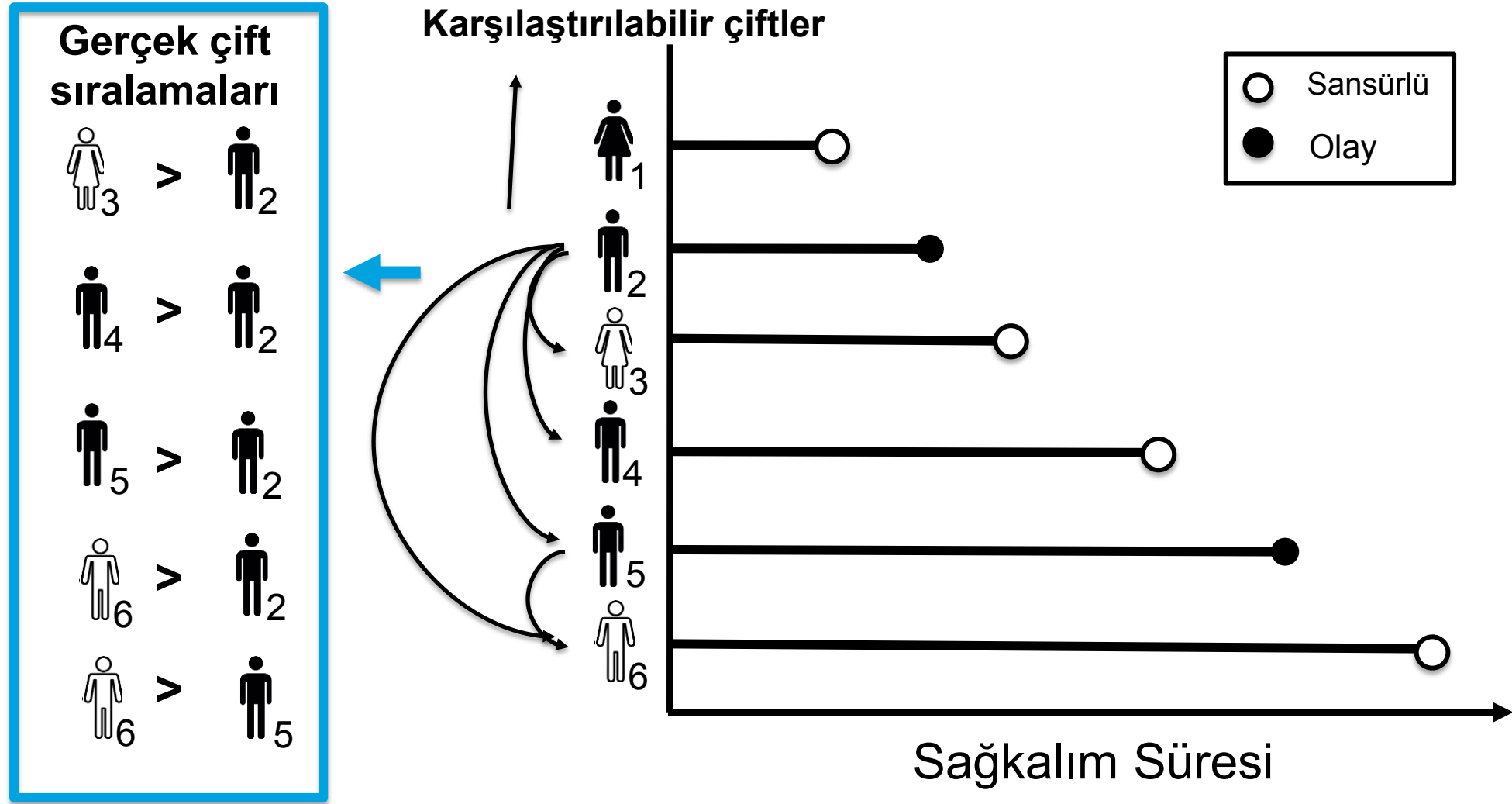
Uyum İndeksi(C-Index)



Uyum İndeksi(C-Index)

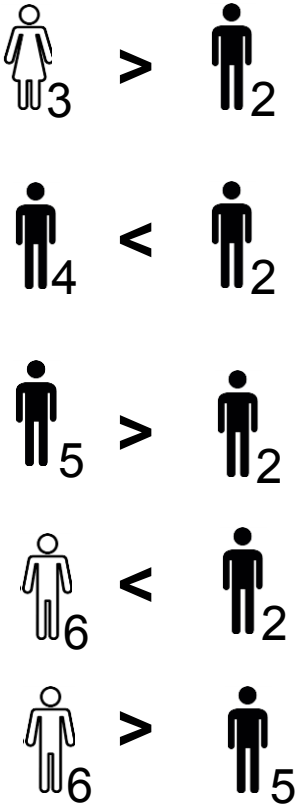


Uyum İndeksi(C-Index)

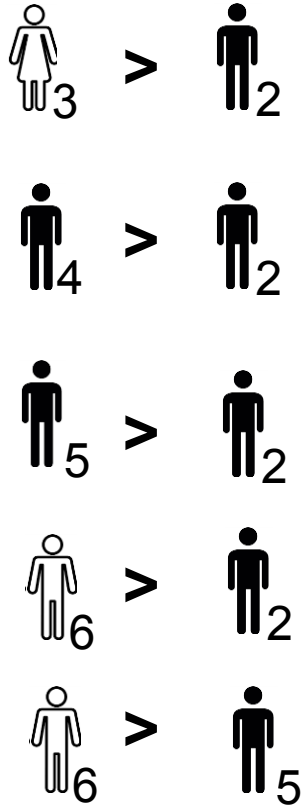


Uyum İndeksi(C-Index)

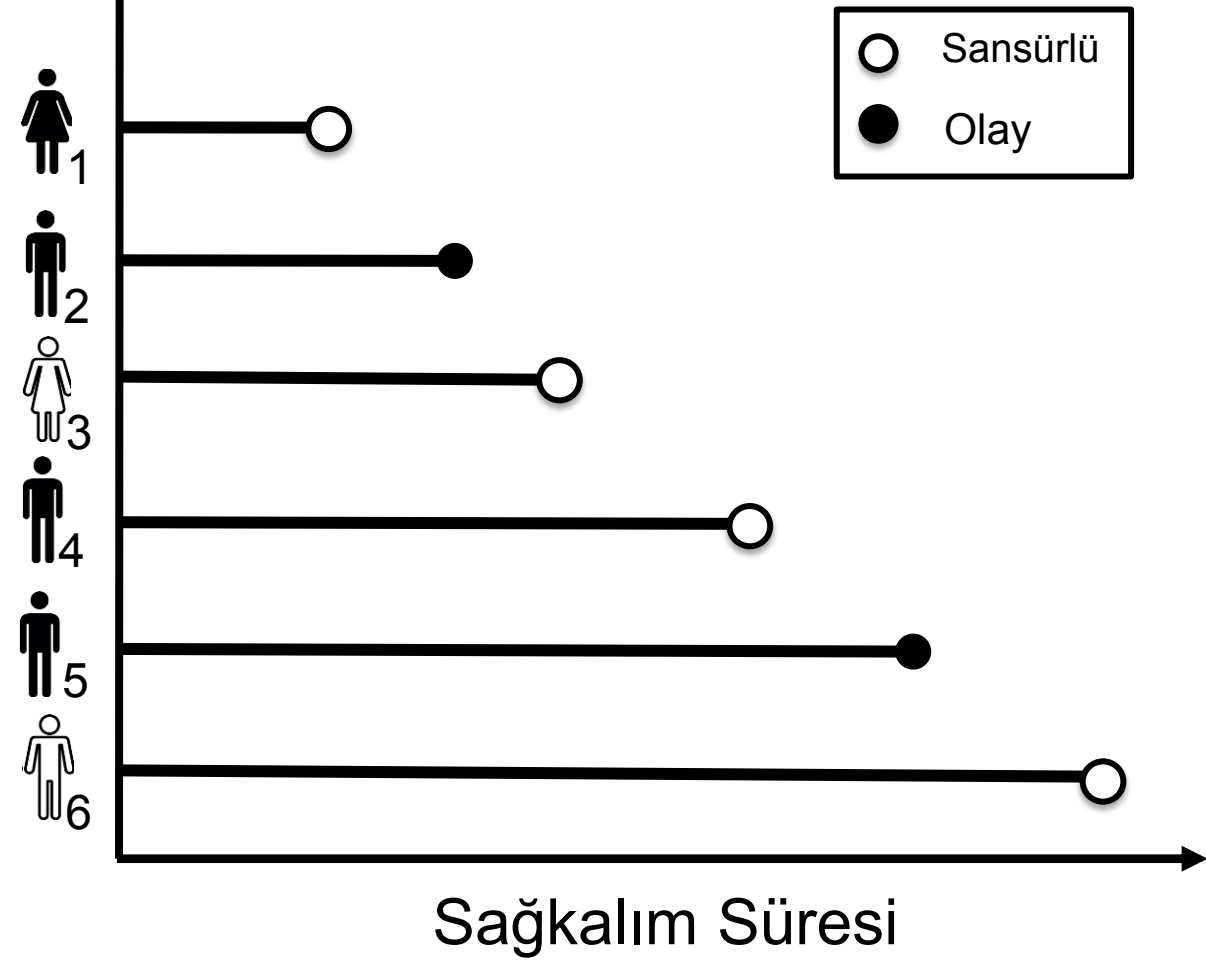
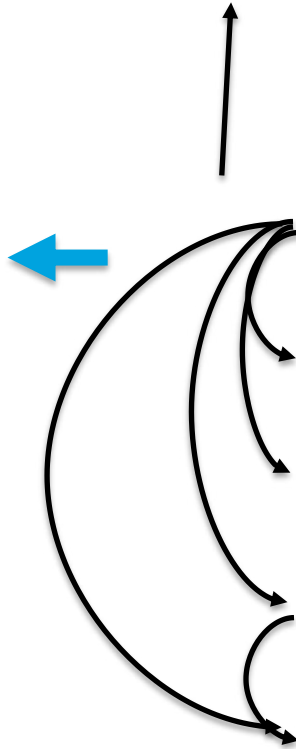
Tahmini çift sıralamaları



Gerçek çift sıralamaları



Karşılaştırılabilir çiftler



Uyum İndeksi(C-Index)

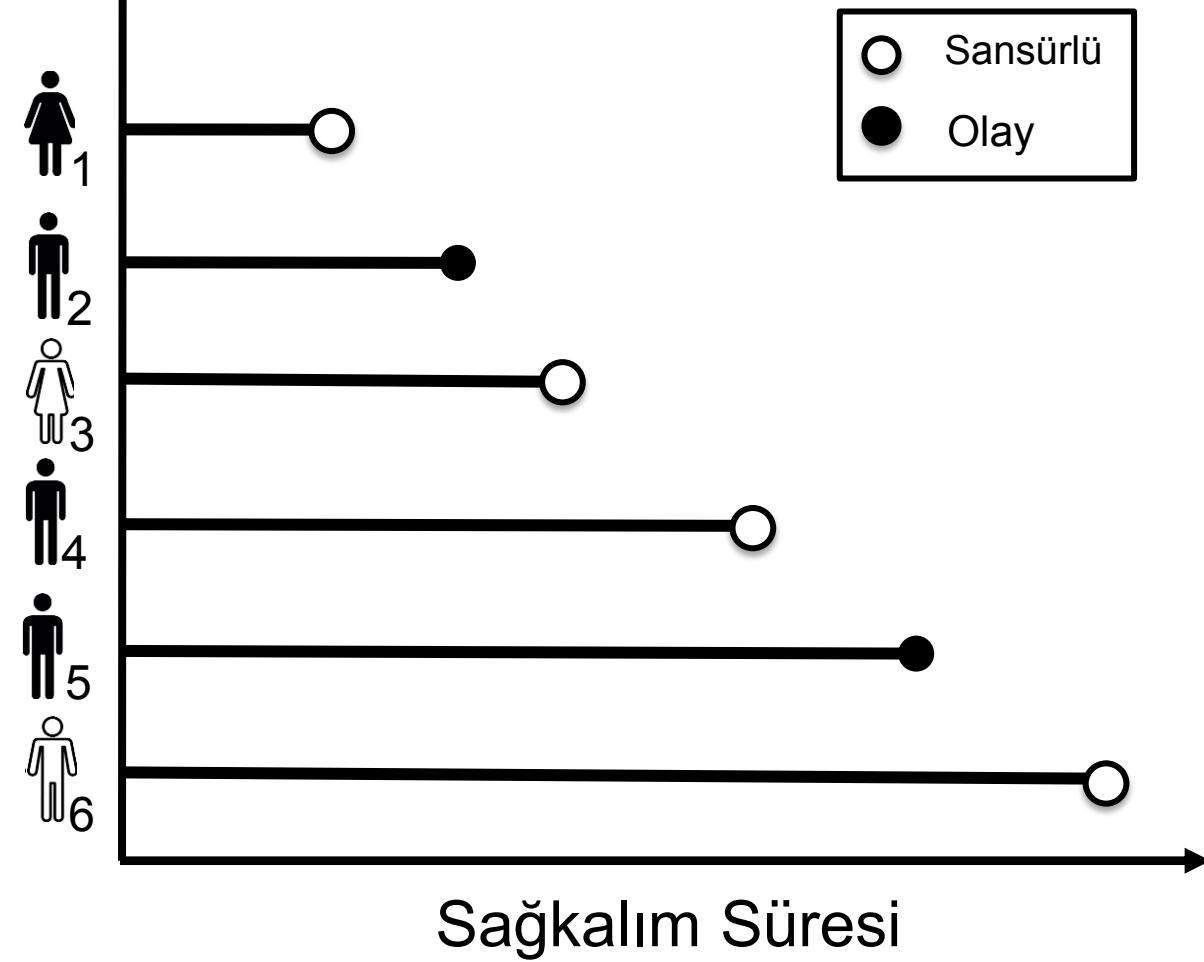
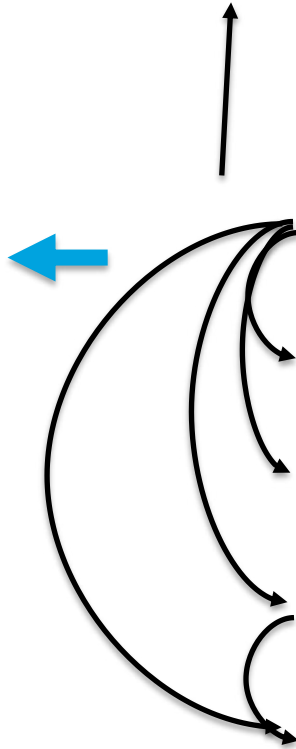
Tahmini çift sıralamaları

	>		✓
	<		✗
	>		✓
	<		✗
	>		✓

Gerçek çift sıralamaları

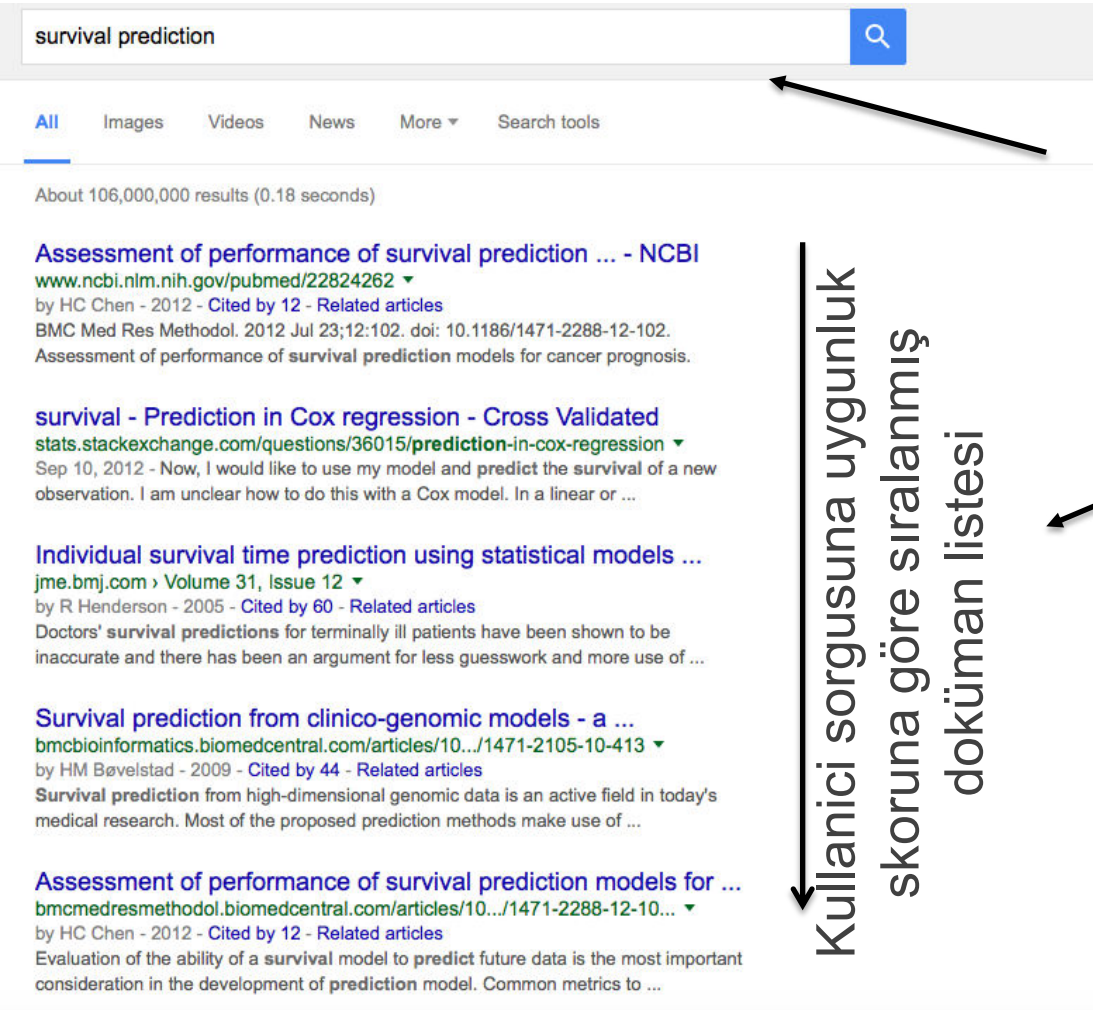
	>		
	>		
	>		
	>		
	>		

Karşılaştırılabilir çiftler



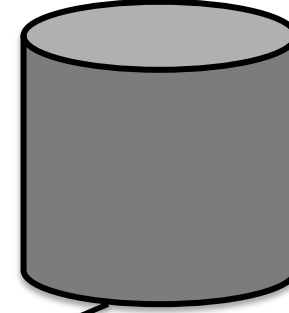
Uyum indeksi = $3/5 = 60\%$

Sıralama Problemi – Webde Arama



The screenshot shows a Google search for 'survival prediction'. The search bar at the top contains the text 'survival prediction' and a magnifying glass icon. Below the search bar, there are tabs for 'All', 'Images', 'Videos', 'News', 'More', and 'Search tools'. The search results are displayed below, showing 'About 106,000,000 results (0.18 seconds)'. The first result is 'Assessment of performance of survival prediction ... - NCBI' with a link to 'www.ncbi.nlm.nih.gov/pubmed/22824262'. The second result is 'survival - Prediction in Cox regression - Cross Validated' with a link to 'stats.stackexchange.com/questions/36015/prediction-in-cox-regression'. The third result is 'Individual survival time prediction using statistical models ...' with a link to 'jme.bmj.com'. The fourth result is 'Survival prediction from clinico-genomic models - a ...' with a link to 'bmcbioinformatics.biomedcentral.com/articles/10.../1471-2105-10-413'. The fifth result is 'Assessment of performance of survival prediction models for ...' with a link to 'bmcmmedresmethodol.biomedcentral.com/articles/10.../1471-2288-12-10...'. A vertical arrow on the right side of the search results points downwards, indicating the flow of the search process.

**Kullanıcı
sorgusu**

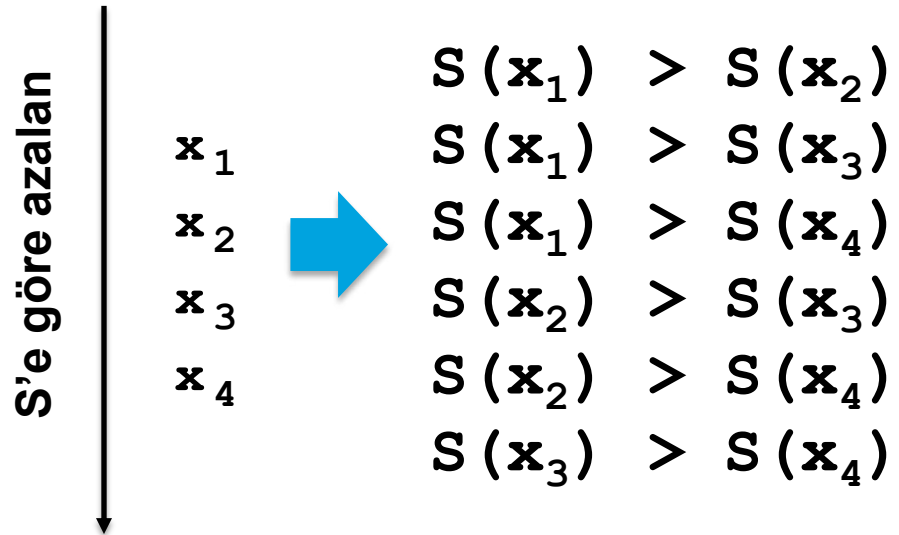


Tüm döküman kümesi
 $D: \{d_1, d_2, \dots, d_N\}$

Kullanıcı sorgusuna uygunluk
skoruna göre sıralanmış
doküman listesi

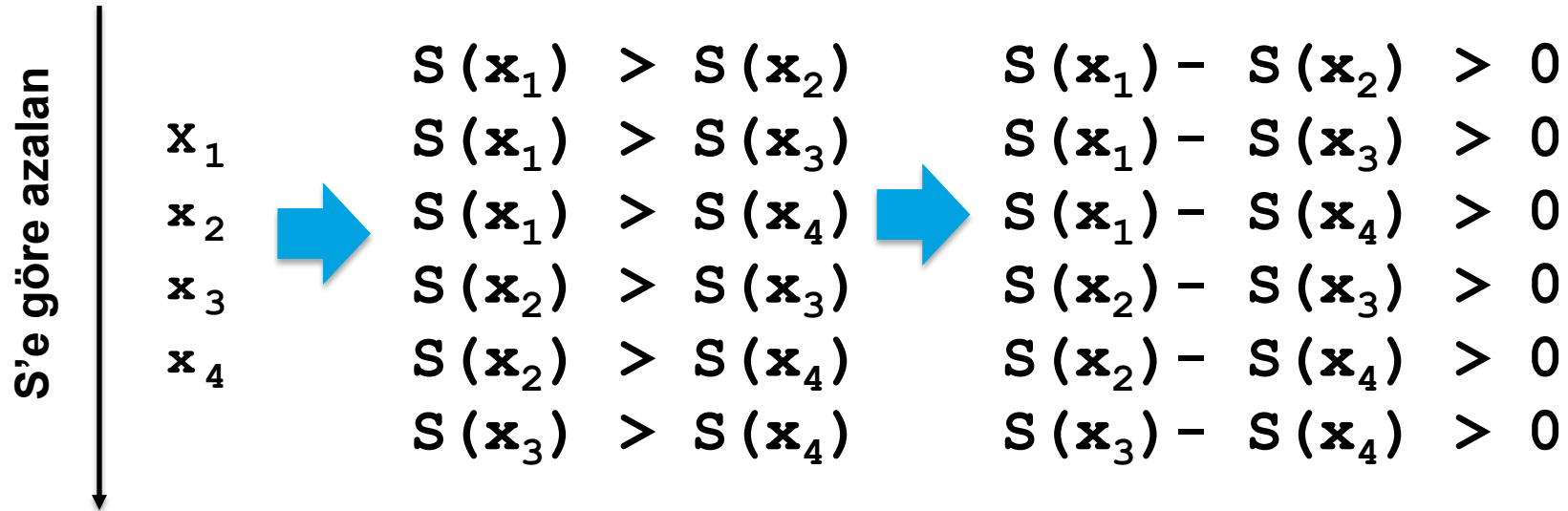
Sıralama Öğrenme – Çift Dönüşümü

Uygunluk skoruna(S) göre sıralanmış n boyutlu bir örneklem listesi, $n(n-1)$ çift sıralamalar ile gösterilebilir.



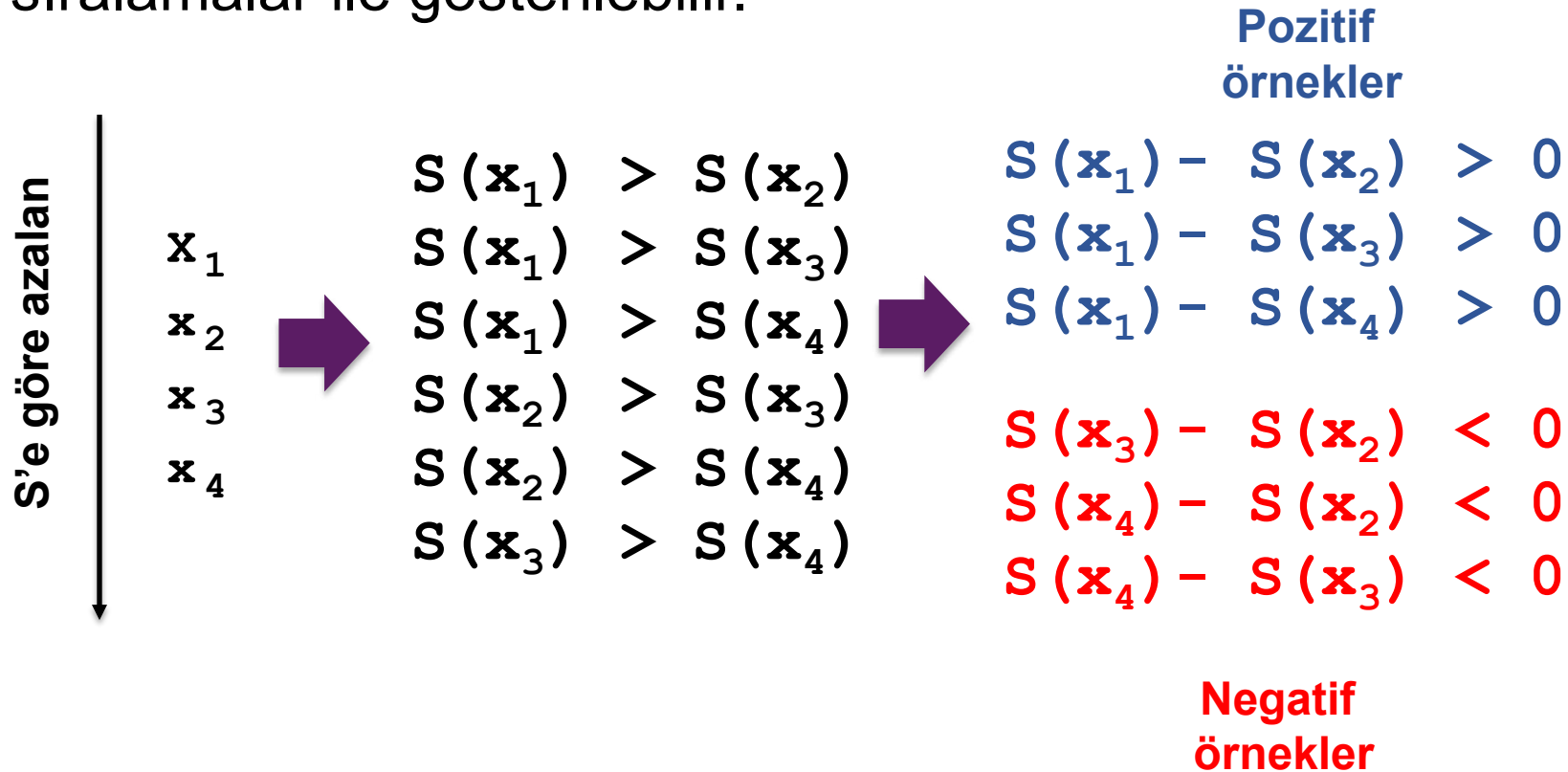
Sıralama Öğrenme – Çift Dönüşümü

Uygunluk skoruna(S) göre sıralanmış n boyutlu bir örneklem listesi, $n(n-1)$ çift sıralamalar ile gösterilebilir.



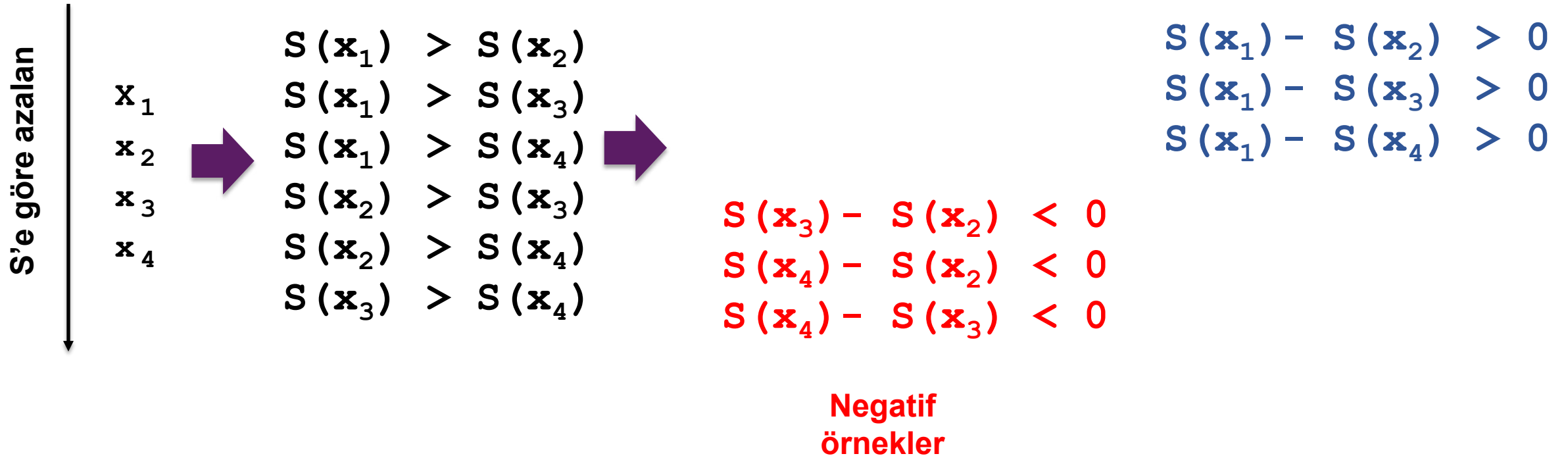
Sıralama Öğrenme – Çift Dönüşümü

Uygunluk skoruna(S) göre sıralanmış n boyutlu bir örneklem listesi, $n(n-1)$ çift sıralamalar ile gösterilebilir.



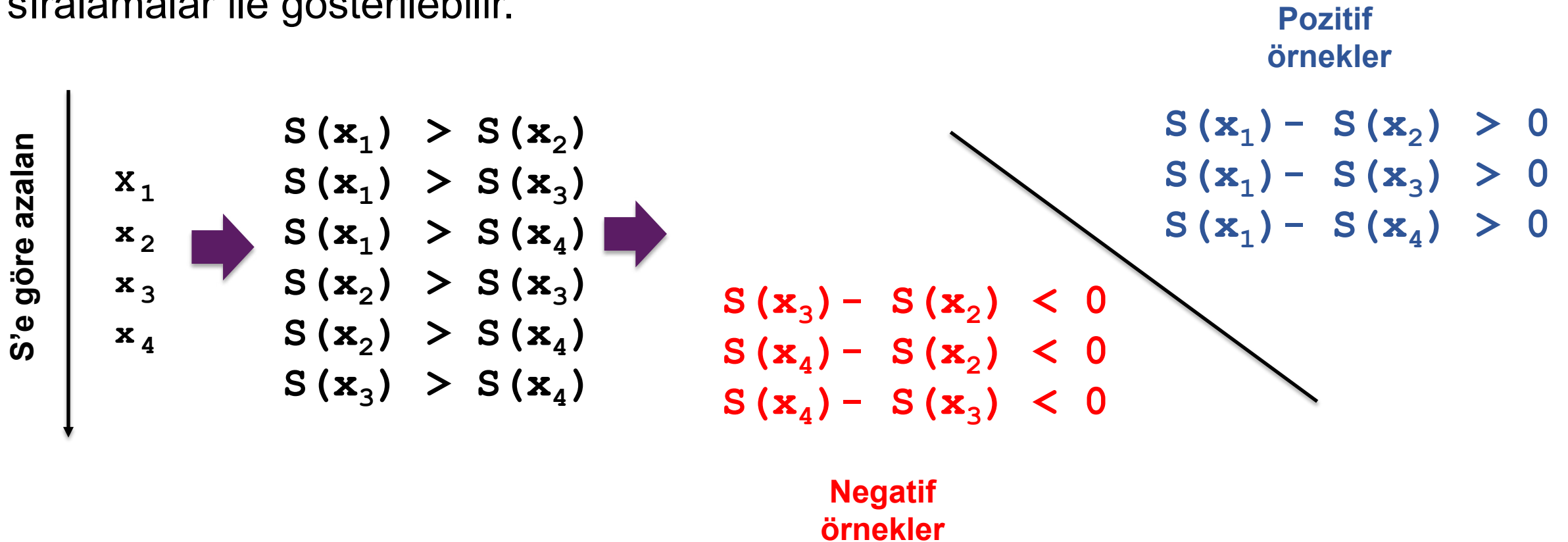
Sıralama Öğrenme – Çift Dönüşümü

Uygunluk skoruna(S) göre sıralanmış n boyutlu bir örneklem listesi, $n(n-1)$ çift sıralamalar ile gösterilebilir.

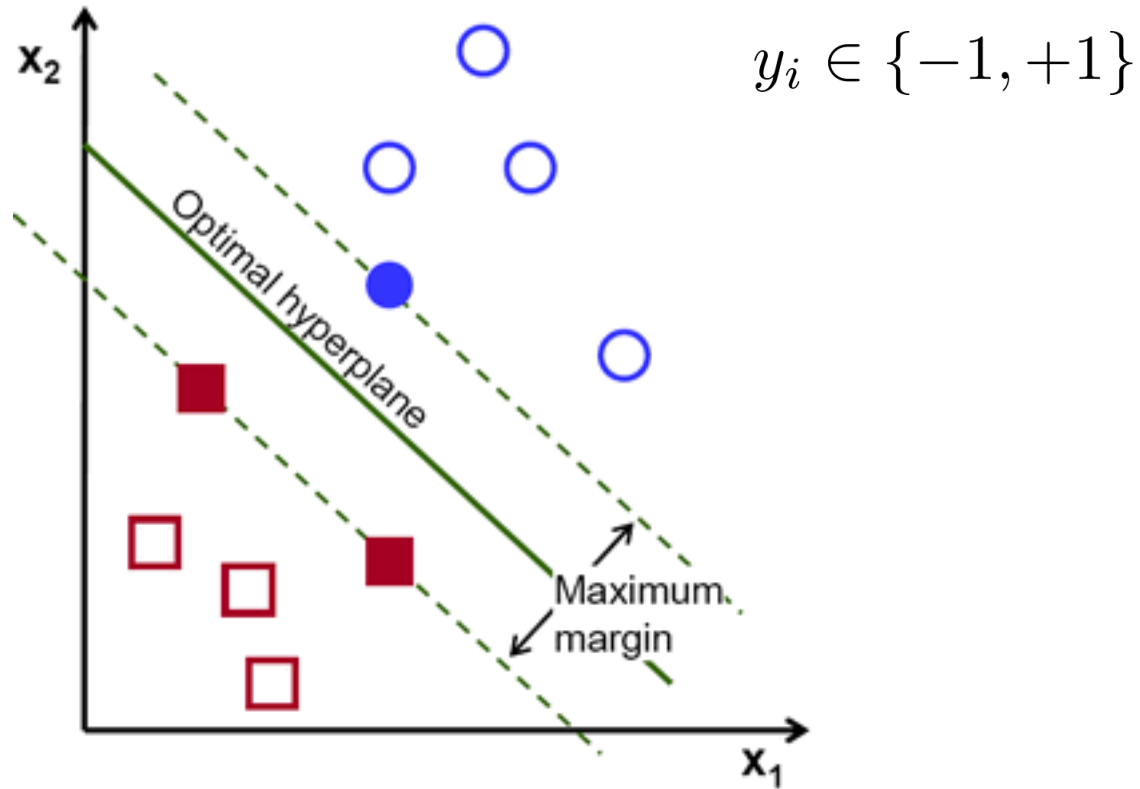


Sıralama Öğrenme – Çift Dönüşümü

Uygunluk skoruna(S) göre sıralanmış n boyutlu bir örneklem listesi, $n(n-1)$ çift sıralamalar ile gösterilebilir.



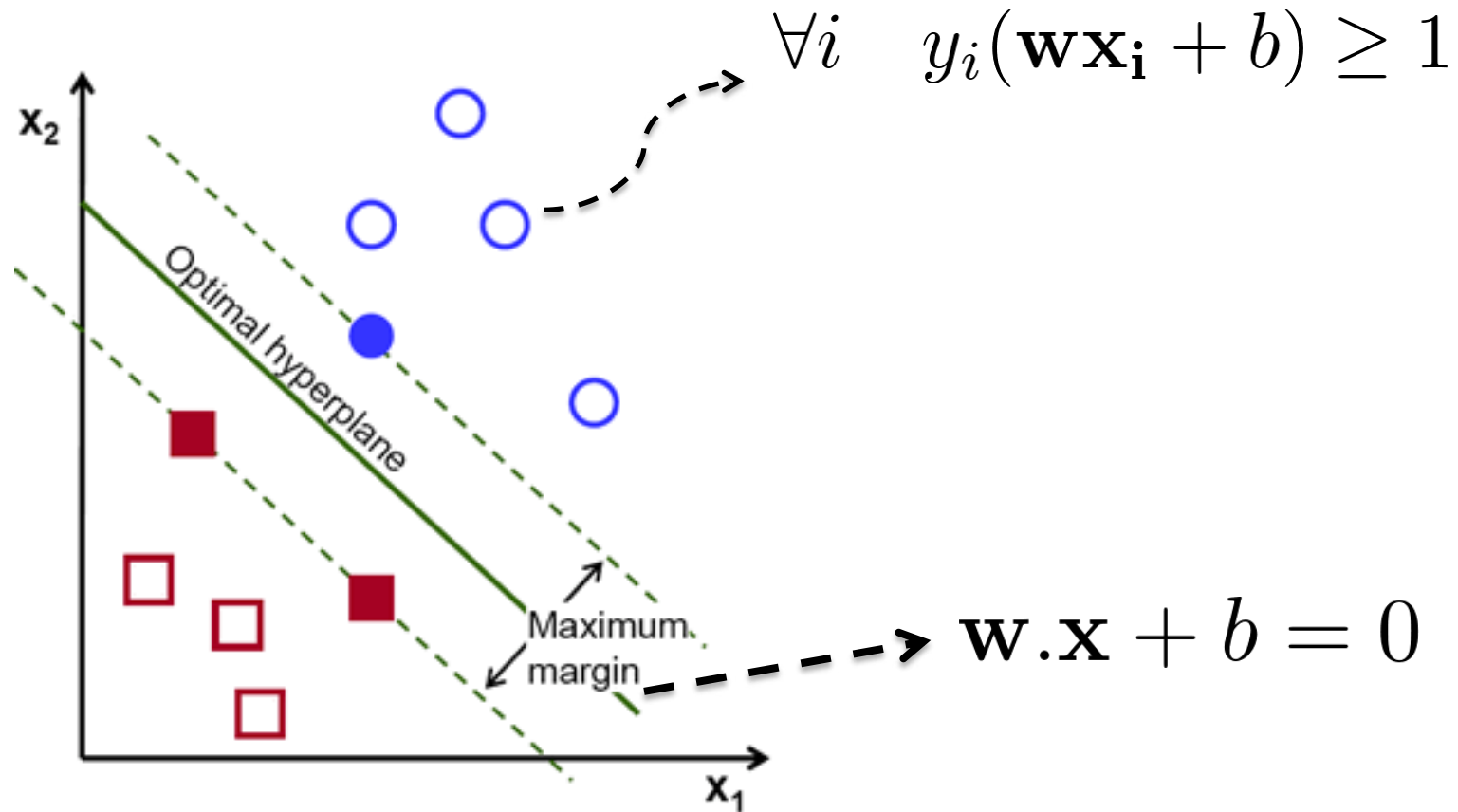
Destek Vektör Makinesi – Sınıflandırma



Cortes, C and Vapnik, V. Machine learning, 20(3), 273–297, 1995.

Figure kaynağı: https://docs.opencv.org/2.4/doc/tutorials/ml/introduction_to_svm/introduction_to_svm.html

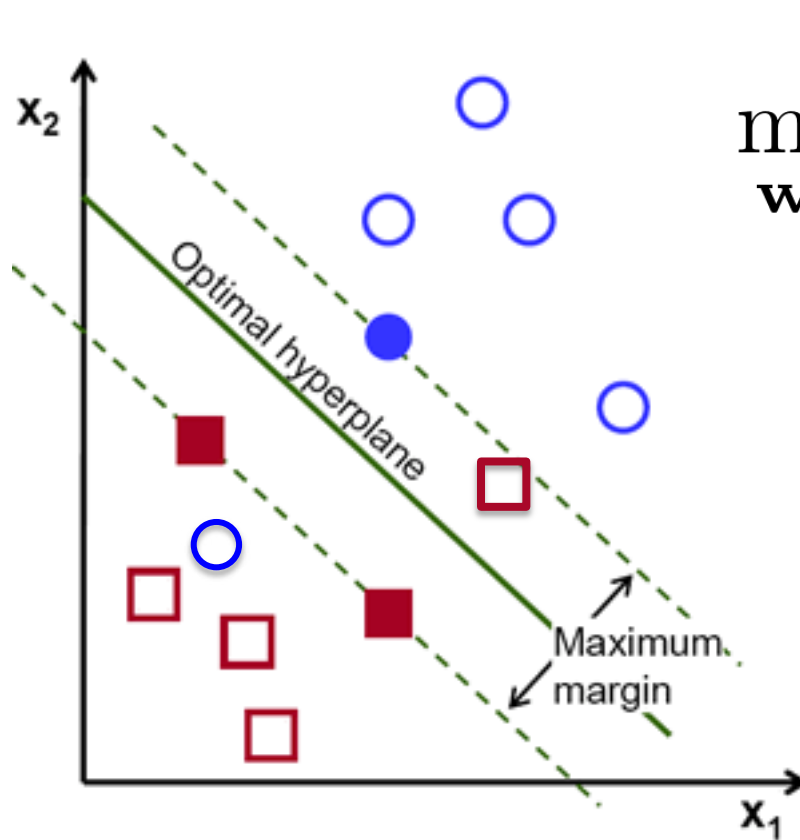
SVM – Sınıflandırma



C. Cortes and V. Vapnik, Machine learning, 20(3), 273–297, 1995.

Figure from: https://docs.opencv.org/2.4/doc/tutorials/ml/introduction_to_svm/introduction_to_svm.html

SVM – Sınıflandırma



Soft-margin case

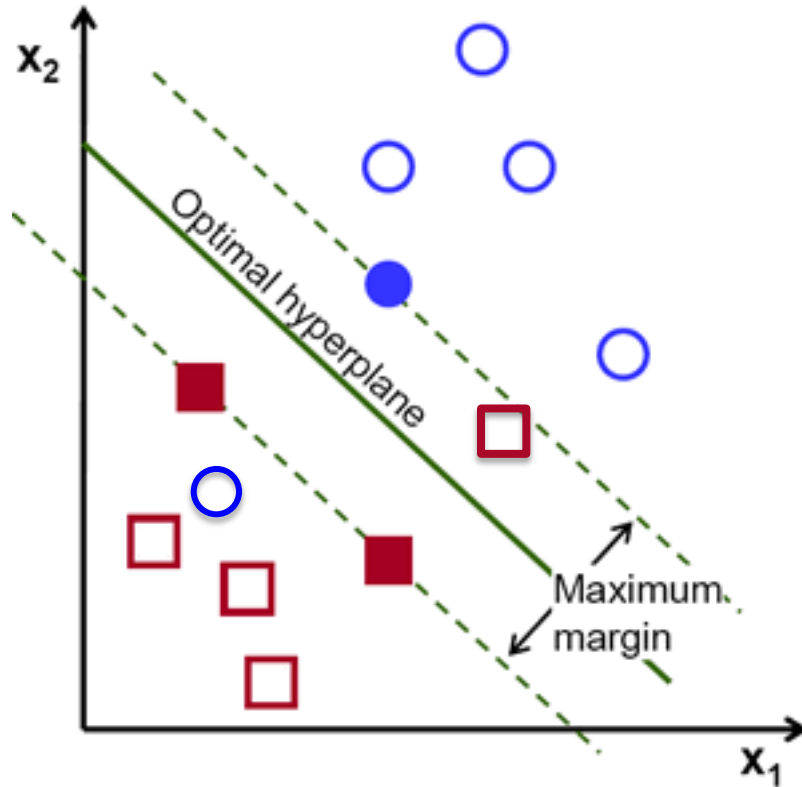
$$\min_{\mathbf{w}, b} \quad \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i$$

$$\text{s.t.} \quad y_i(\mathbf{w}x_i + b) \geq 1 - \xi_i$$

$$\xi_i \geq 0$$

Örnekler hiperdüzlemin doğru tarafında yer almalı

RankSVM - Sıralama



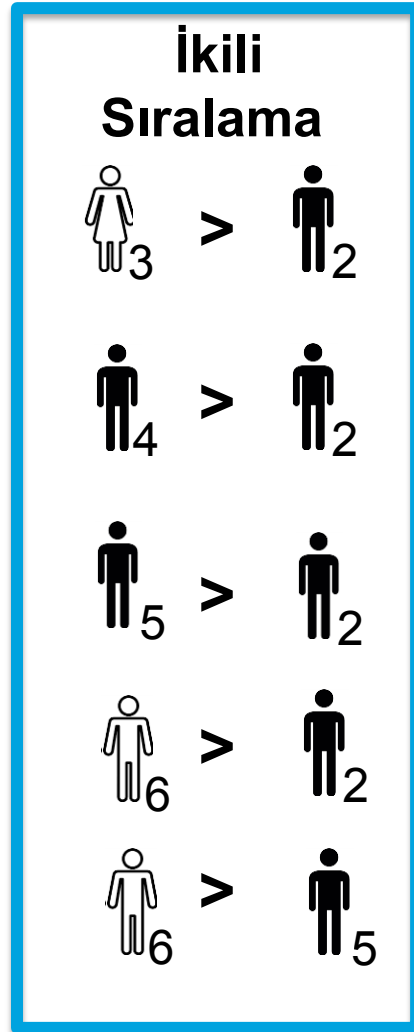
$$\min_{\mathbf{w}, b} \quad \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_{i,j}$$

$$\text{s.t.} \quad \{\forall(i, j) \quad S(x_i) > S(x_j) :$$

$$\mathbf{w} \cdot (\mathbf{x}_i - \mathbf{x}_j) \geq 1 - \xi_{i,j}$$

$$\forall(i, j) \quad \xi_{i,j} \geq 0$$

RsurVM – Kısıtları Modifiye Et

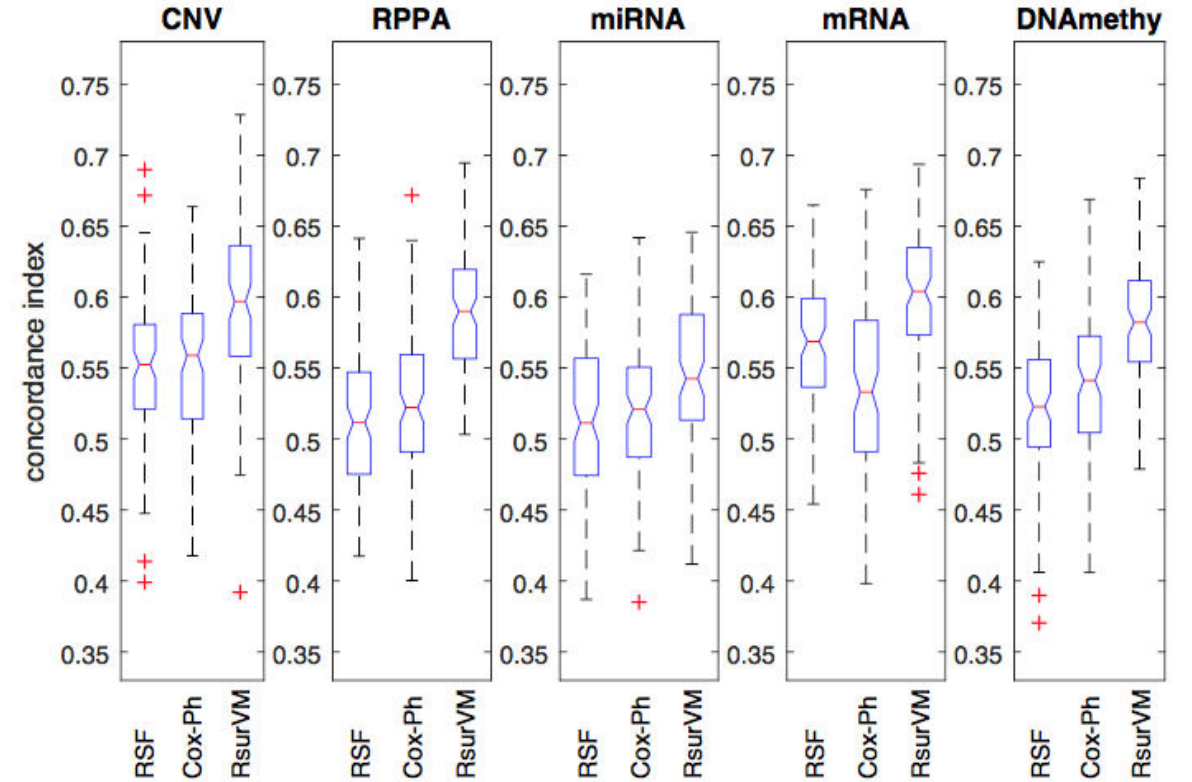
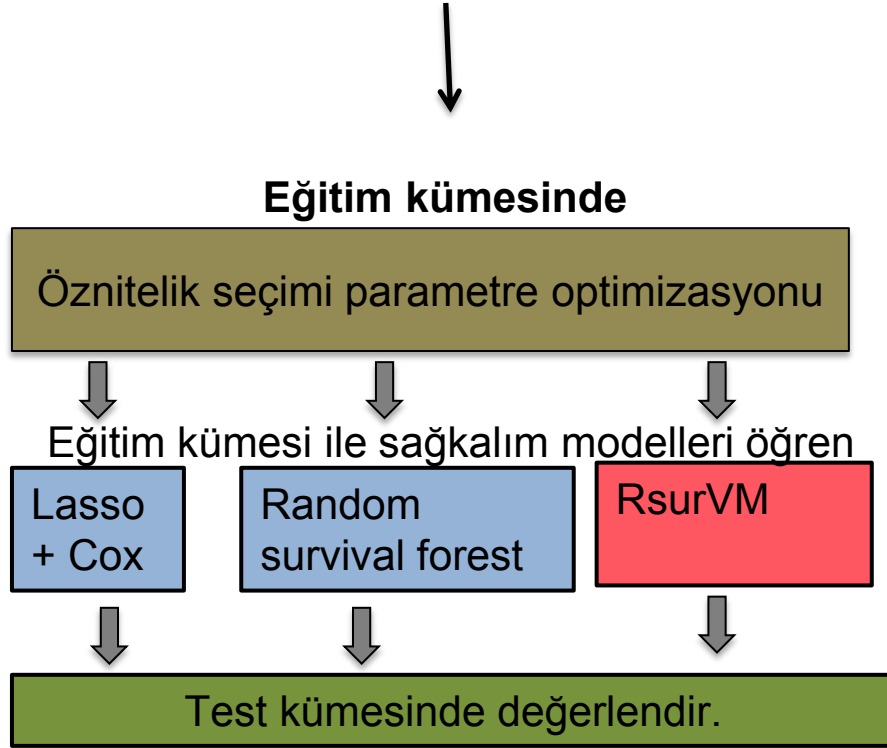


Hasta çiftlerinin yaşam sürelerine göre kısıtlar ekle ve sınıflandırma problemini çöz

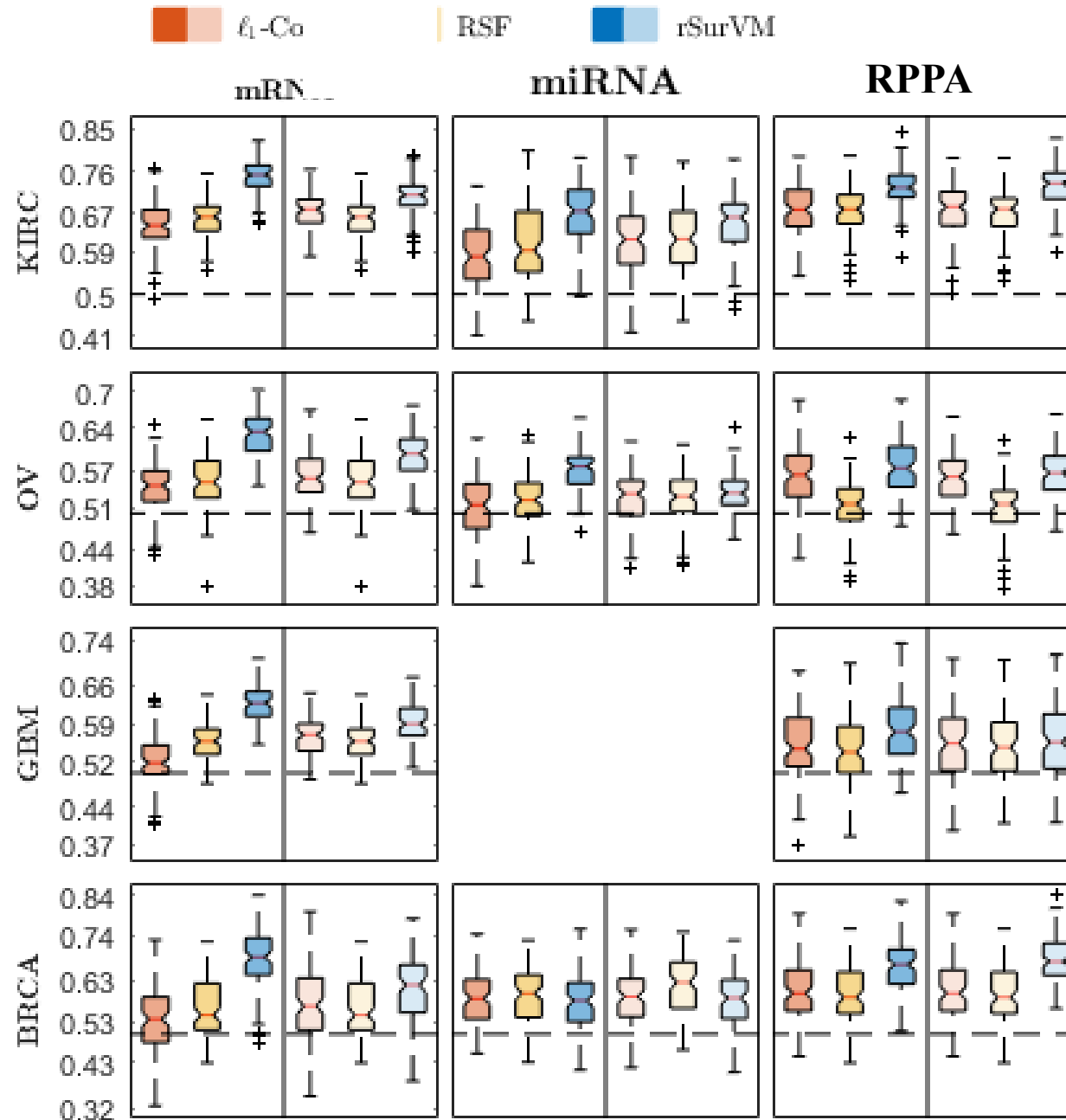
$$S(\mathbf{x}_i) > S(\mathbf{x}_j) \Rightarrow \mathbf{w} \cdot (\mathbf{x}_i - \mathbf{x}_j) > 0$$

DeneySEL DÜZENEK

100 kez rastgele eğitim ve test kümesi ayır

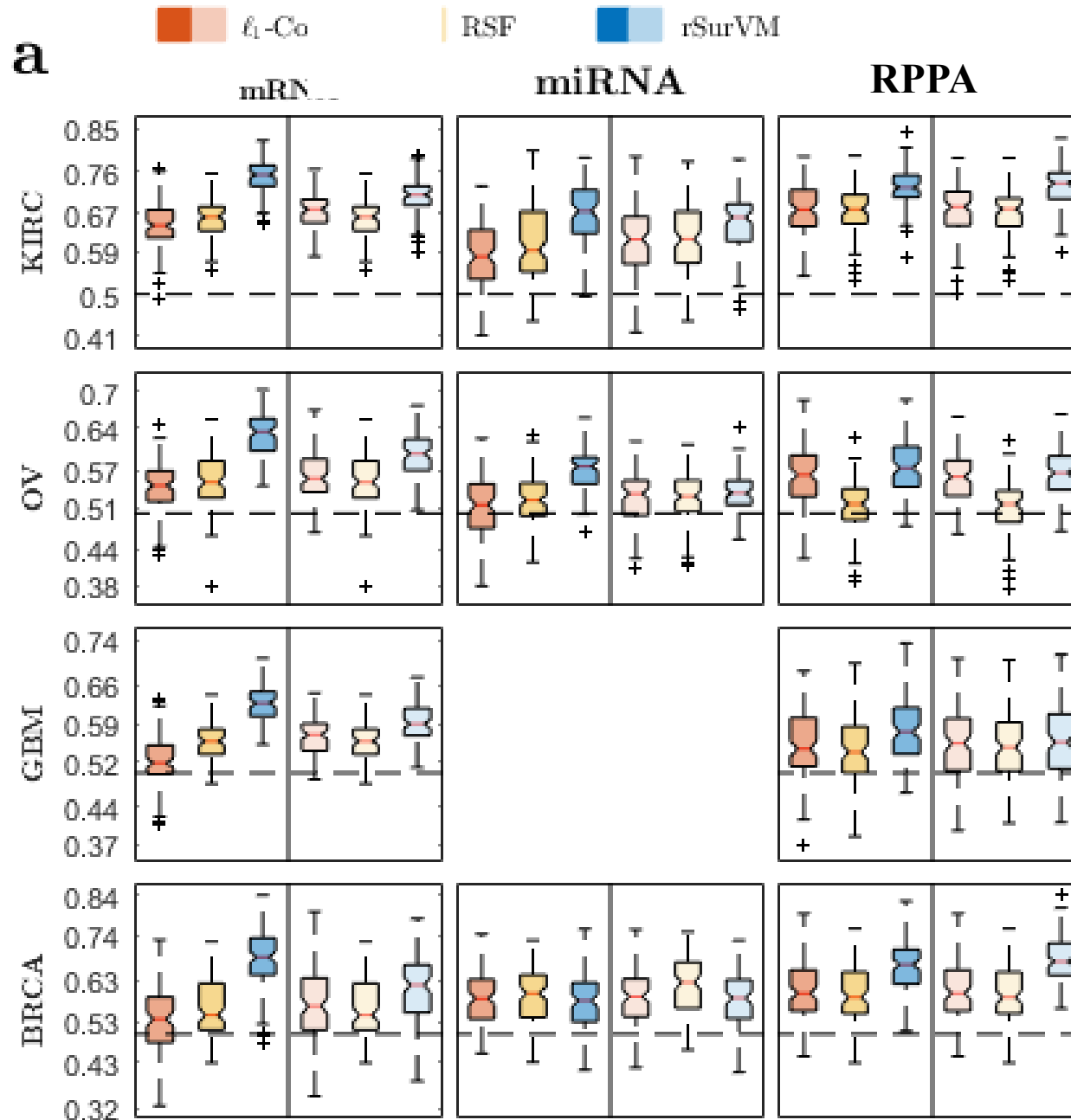


RsurVM kullanılan moleküler veri tipinden bağımsız olarak, şu an kullanılan en geçerli yöntemlerden daha iyi tahminliyor.



Koyu renk
Öznitelik filtresi olmaksızın.

Açık renk:
Uni-cox filtresi



Koyu renk
Öznitelik filtresi olmaksızın.

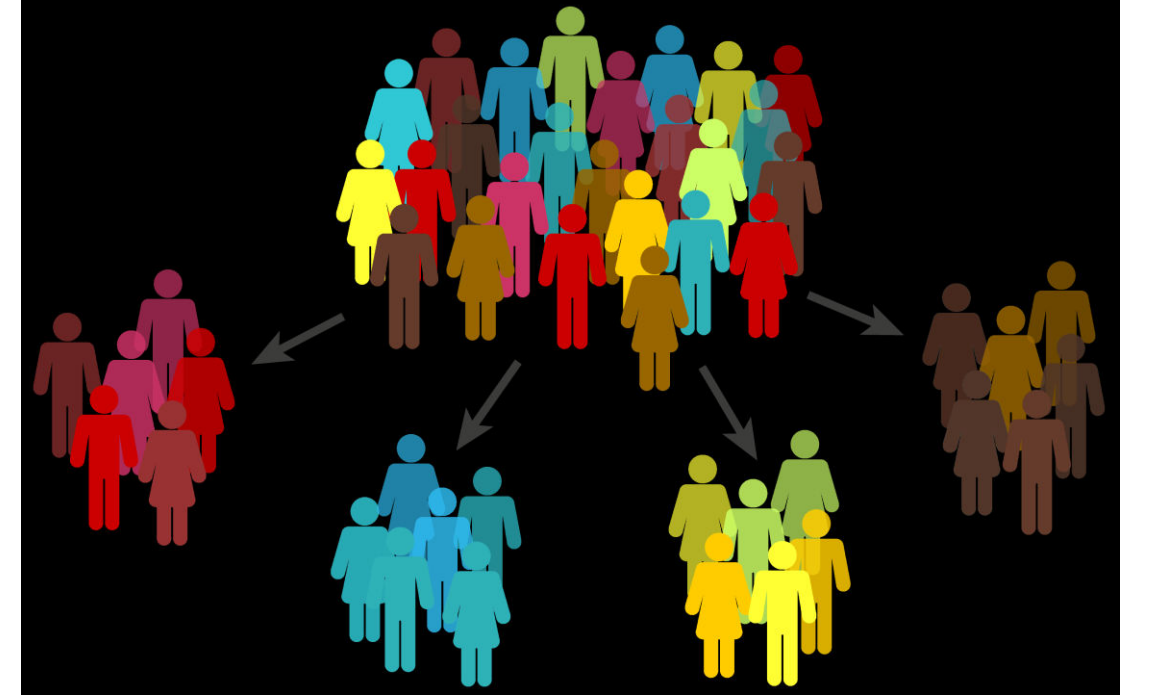
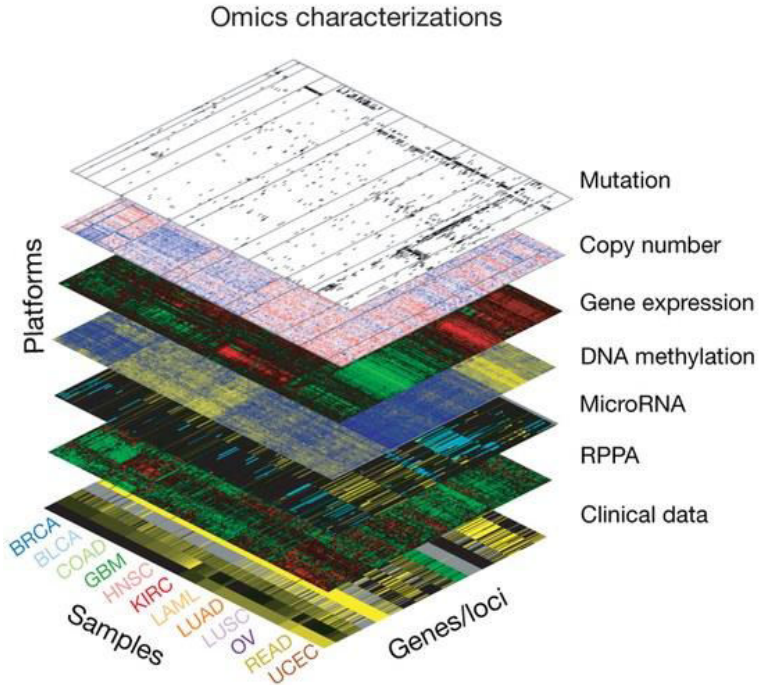
Açık renk: Uni-cox filtresi



Mustafa Büyüközkan
Önceden, Bilkent Üniversitesi
Şimdi PhD Helmholtz Zentrum
München

Kompleks Hastalıklarda Hasta Alt-Gruplarının Tanımlamaya Yönelik Makine Öğrenme Tekniklerinin Geliştirilmesi

- Kompleks hastalıklar hetorejendir
- Tedavi ve tanı için alt grupların keşfi



Yarı-Gözetimli Öbekleme ile Hasta Gruplarının Bulmak

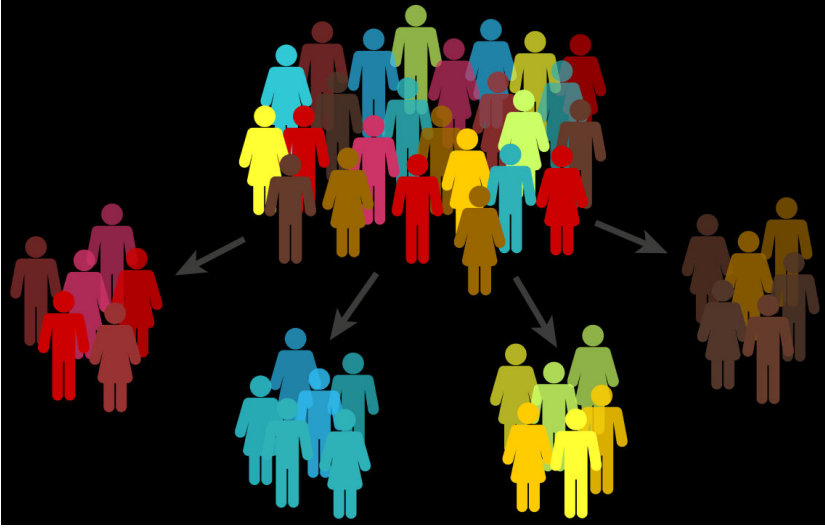
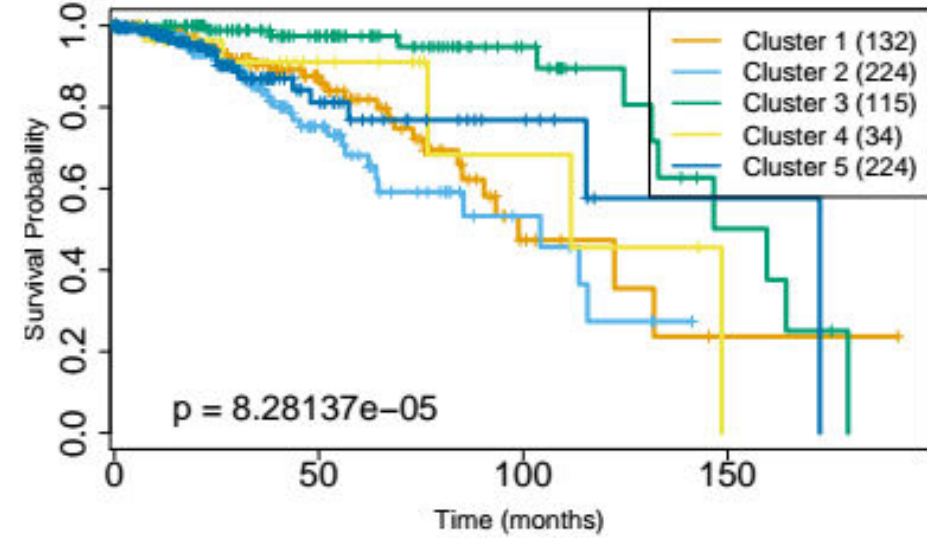


Figure kaynağı

<http://www.pancreaticcancer.net.au/ground-breaking-nature/>



Duygu Ozcelik
Bilkent Üniversitesi
Şimdi, Havelsan



Yarı-Gözetimli Öbekleme ile Hasta Gruplarını Bulmak

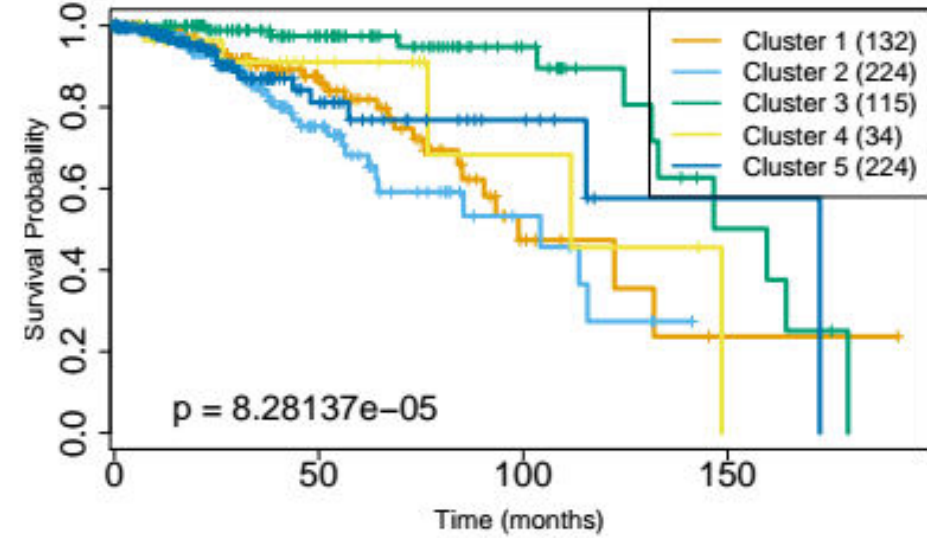
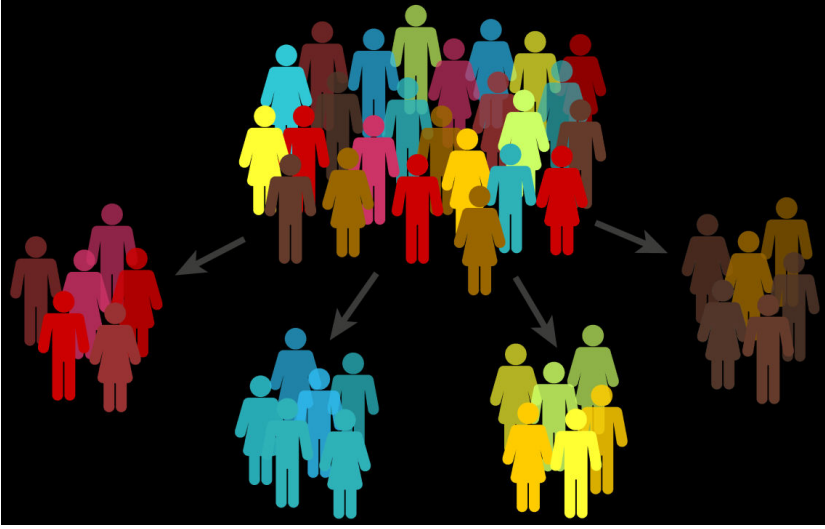


Figure kaynağı
<http://www.pancreaticcancer.net.au/ground-breaking-nature/>

Yarı-Gözetimli Öbekleme ile Hasta Gruplarını Bulmak

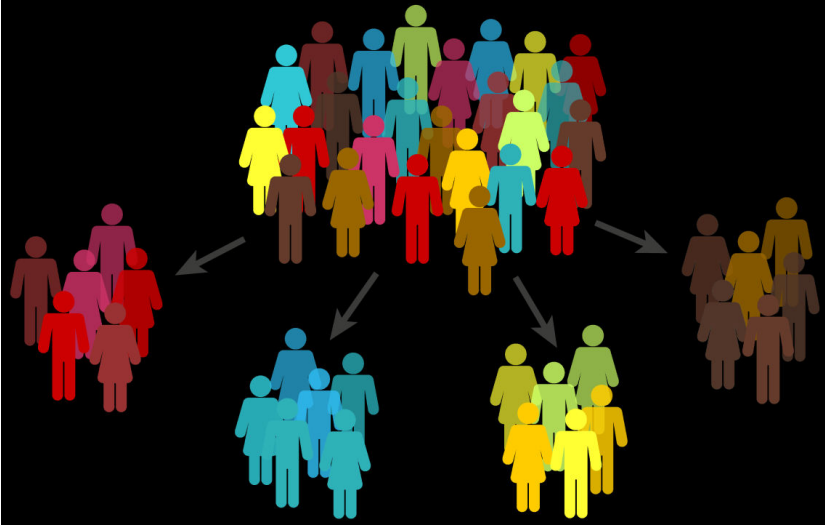
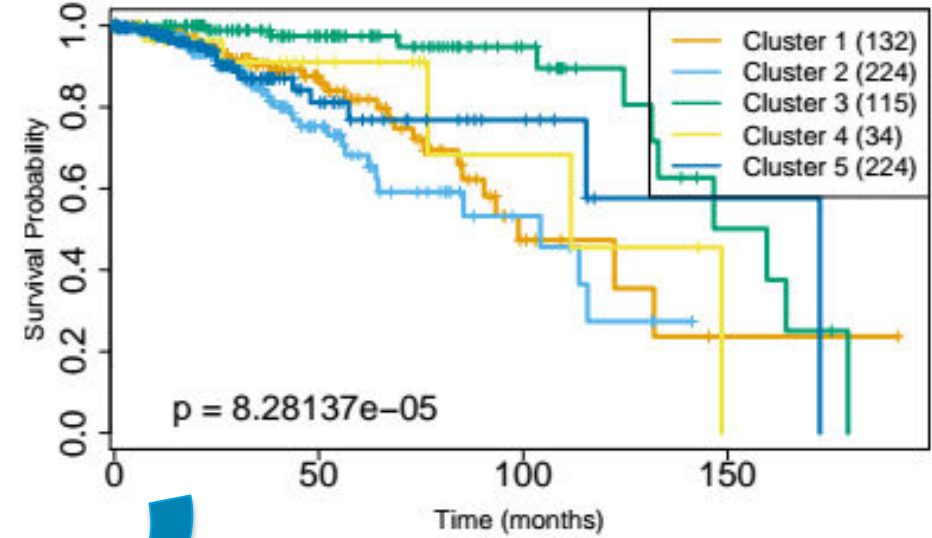
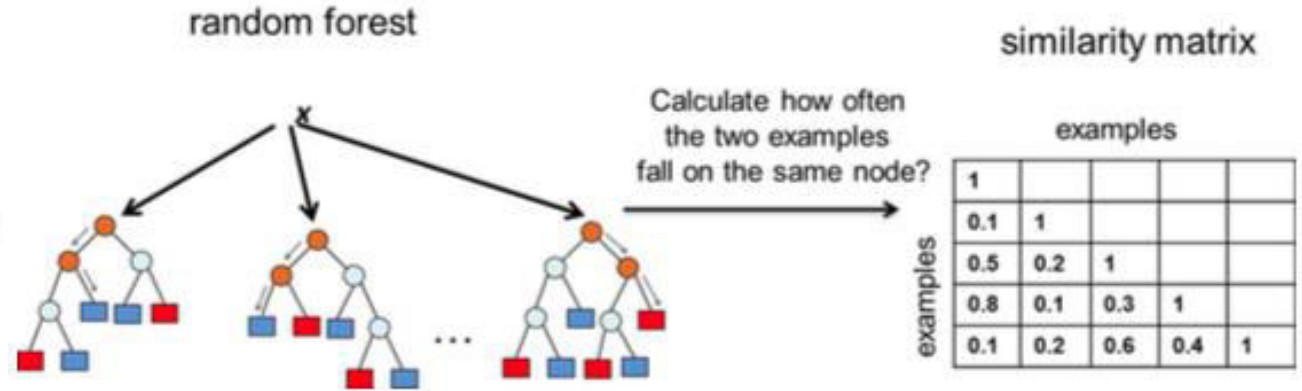
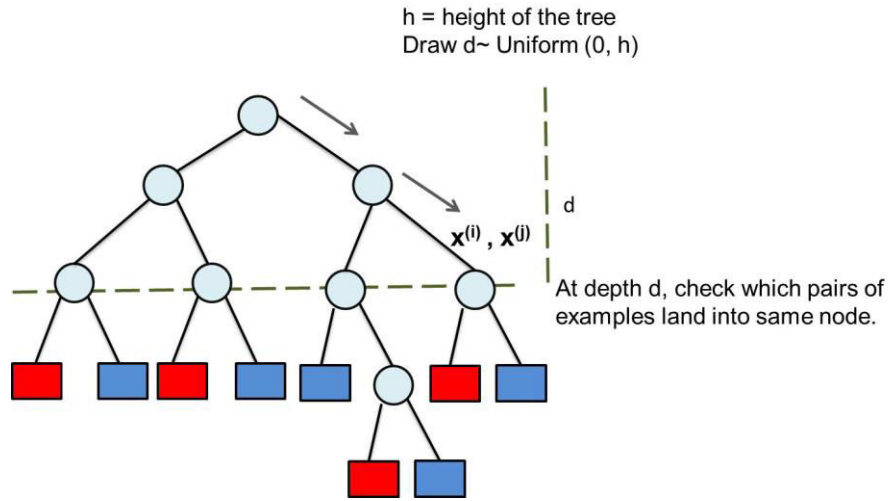


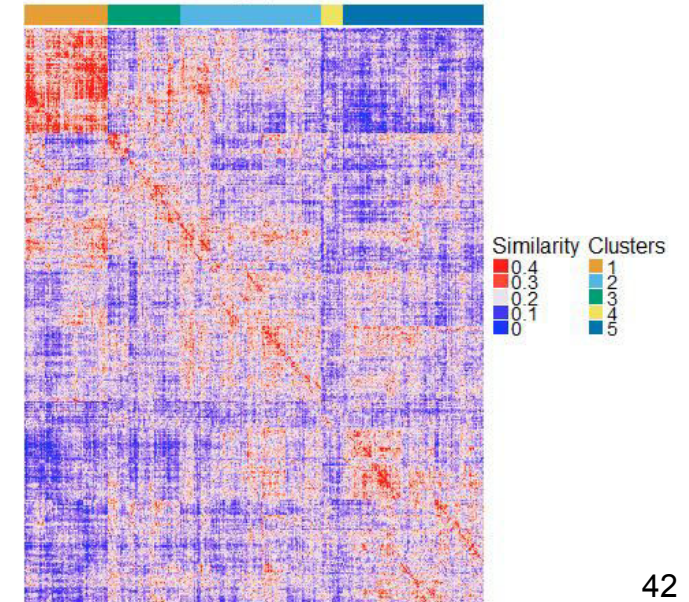
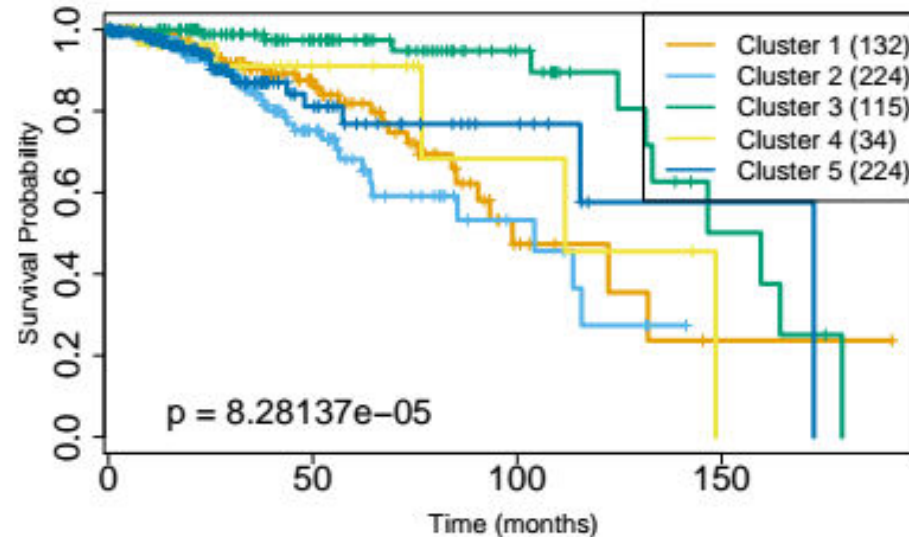
Figure kaynağı
<http://www.pancreaticcancer.net.au/ground-breaking-nature/>



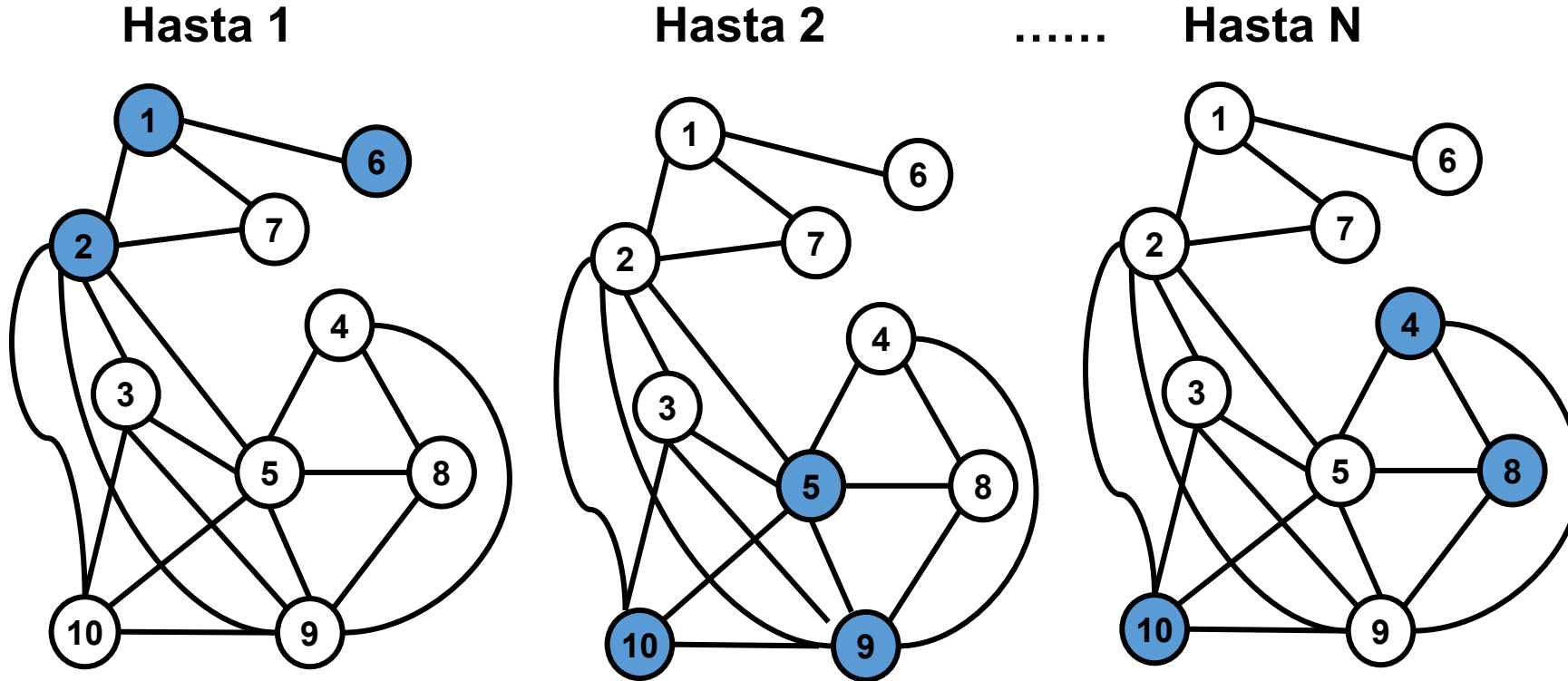
Yarı-Gözetimli Öbekleme ile Hasta Altgruplarının Bulunması



Duygu Ozcelik
Bilkent Üniversitesi
Şimdi, Havelsan



Hasta Benzerliklerini Yolak Çizge Çekirdekleri ile Hesaplama

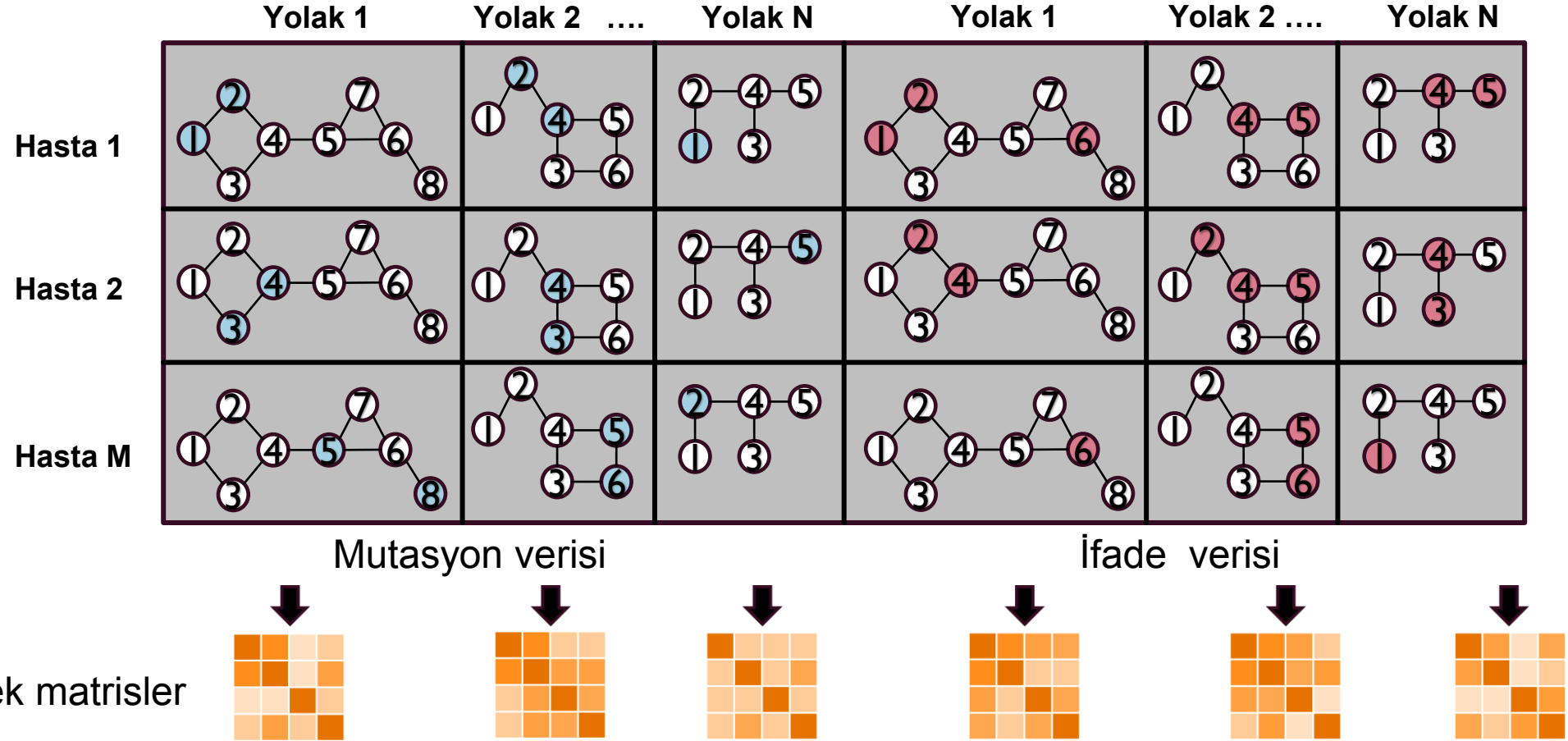


Önceden,
Bilkent Üniversitesi

Şimdi, PhD, Tübingen
Üniversitesi

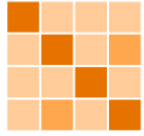
Düğümün gen
Renk hastadaki değişim (mutasyon, ifade değişikliği)

Birden çok yolak, birden çok veri tipi

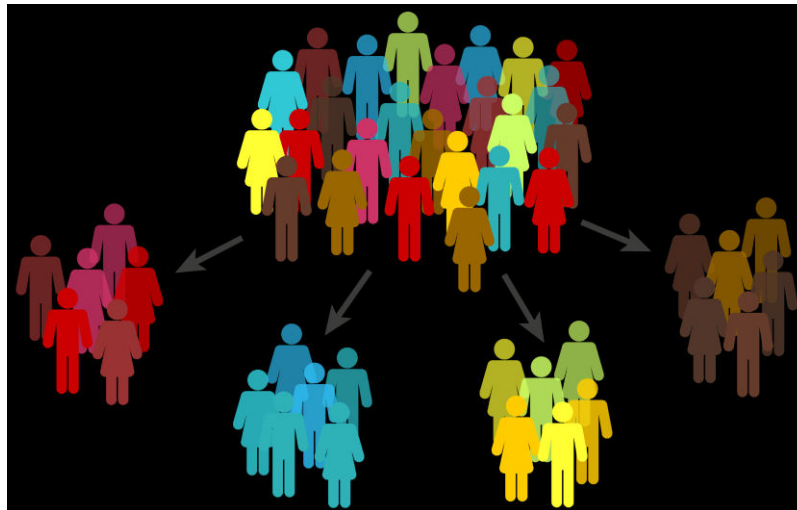


Çekirdekli Çok Yönlü Kümele

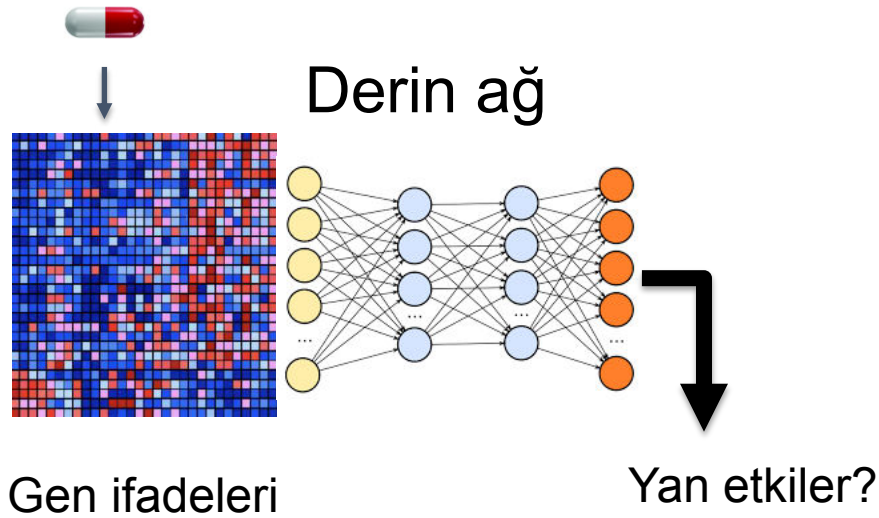
Çizgiler ve değişimler üzerinden hesaplanmış matris



Çok Yönlü Çekirdekli Kümele



İlaç yan etkilerinin derin öğrenme ile tahmini



Onurcan Üner

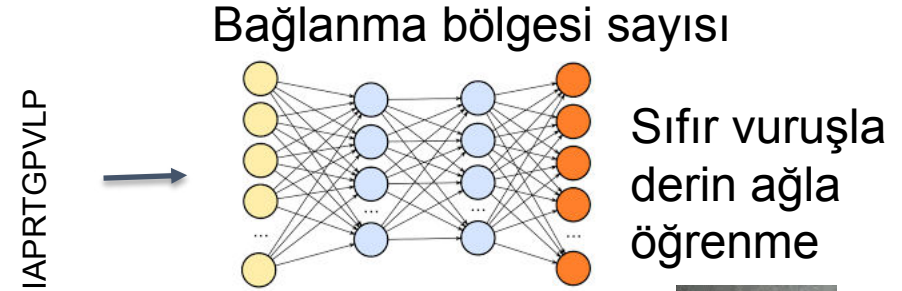
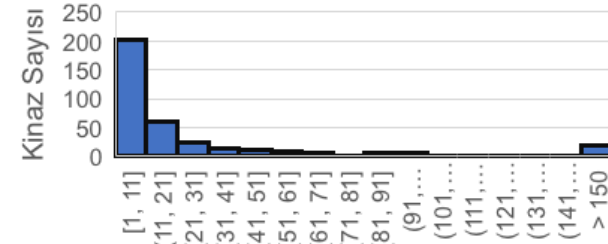


Dr. Gokberk
Cinbis



Dr. Ercüment
Çiçek

Fosforilasyon bölgesinden bağlanan kinazı tahmin etme



İman Deznaby

Dr. Mehmet
Koyutürk



SPADIS: Genom Boyu İlişkilendirme Çalışmalarında Çeşitleyici ve Açıklayıcı SNP küme seçimi

Bir fenotiple ilişkili bir SNP setini seçmek
Makine öğrenme bağlamında özellik alt kümesi seçim problemi

Zorluklar:

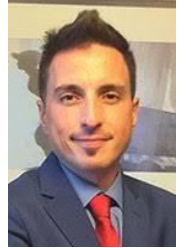
- SNP'ler tek tek fenotip ile ilişkili değildir
- Birden çok SNP'nin ortak etkisi
- Düşük örnek büyüklüğü (10^2 - 10^4 kişi)
- Yüksek boyut (10^5 - 10^7 SNP)

Yaklaşım:

- Bir ön bilgi ile oluşturulmuş SNP-SNP ağı kullanarak işlevsel etkileşimleri hesaba katmak.

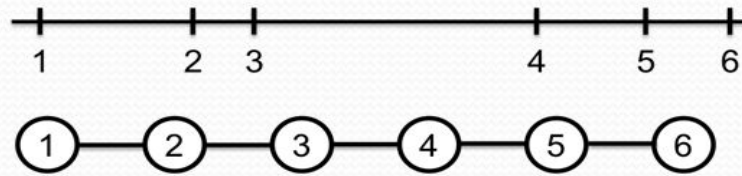


Serhan
Yılmaz



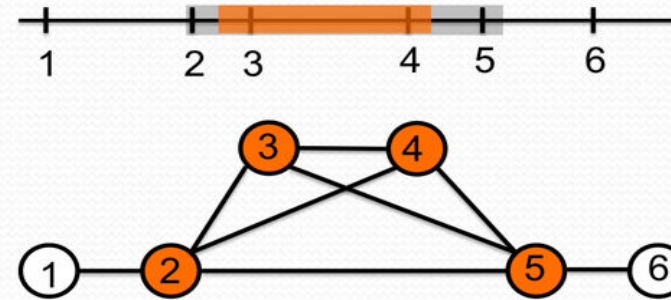
Dr.
Ercüment
Çiçek

Gene Sequence (GS)



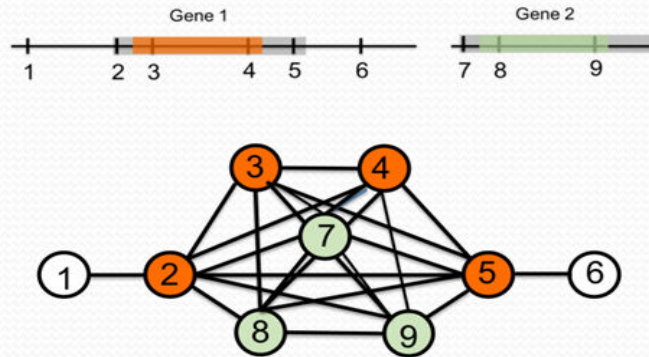
- SNPs adjacent on the sequence

Gene Membership (GM)



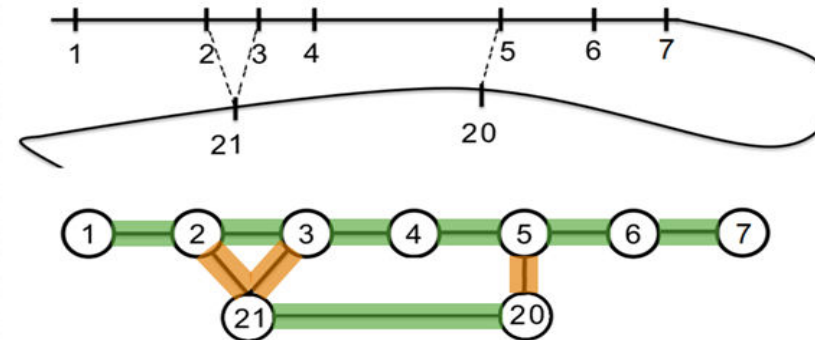
- SNPs adjacent on the sequence
- SNPs near the same gene

Gene Interaction (GI)



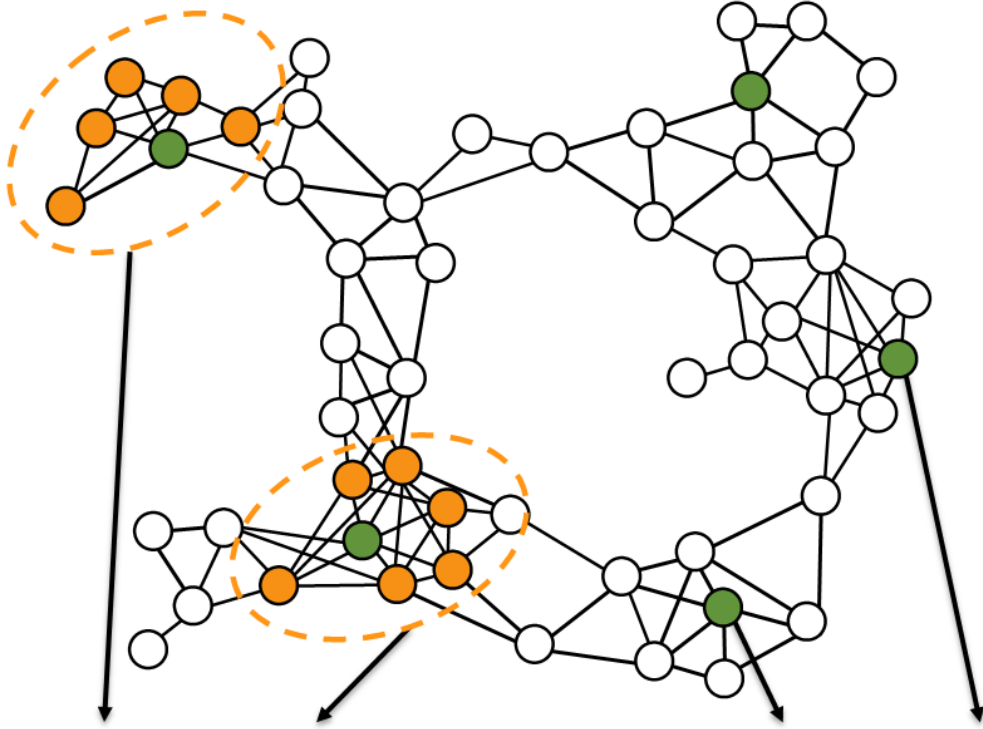
- SNPs adjacent on the sequence
- SNPs near the same gene
- SNPs near interacting genes

Gene Sequence 3D (GS-HICN)



- SNPs adjacent on the sequence
- SNPs significantly close in 3D

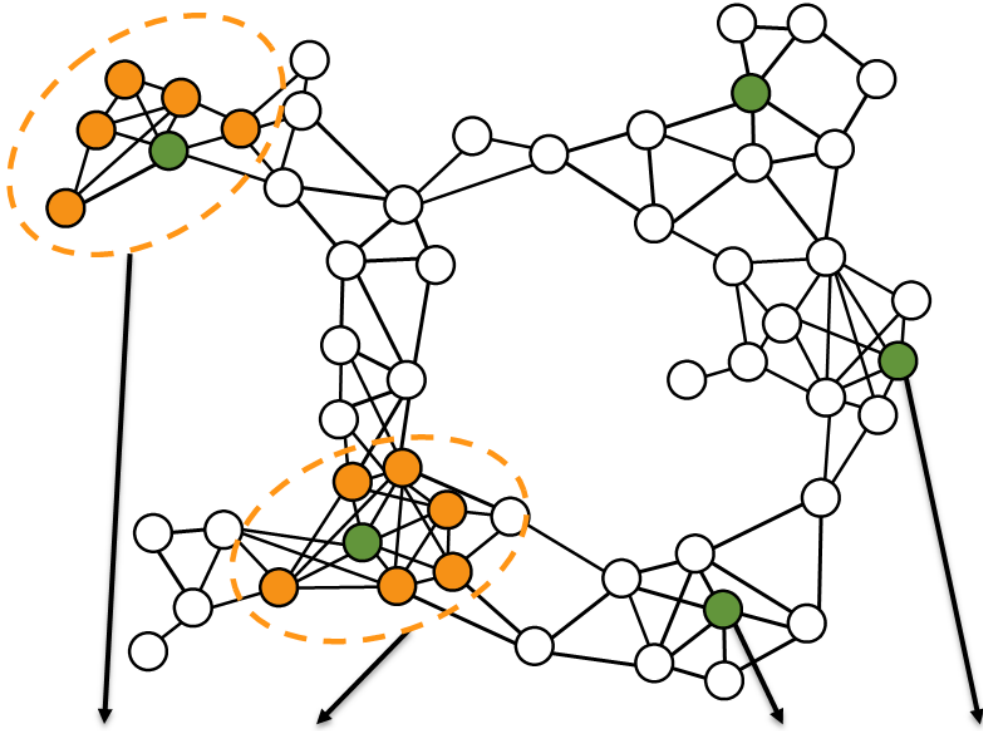
SPADIS



Yakın SNPler tekrarlayan bilgi getirir.

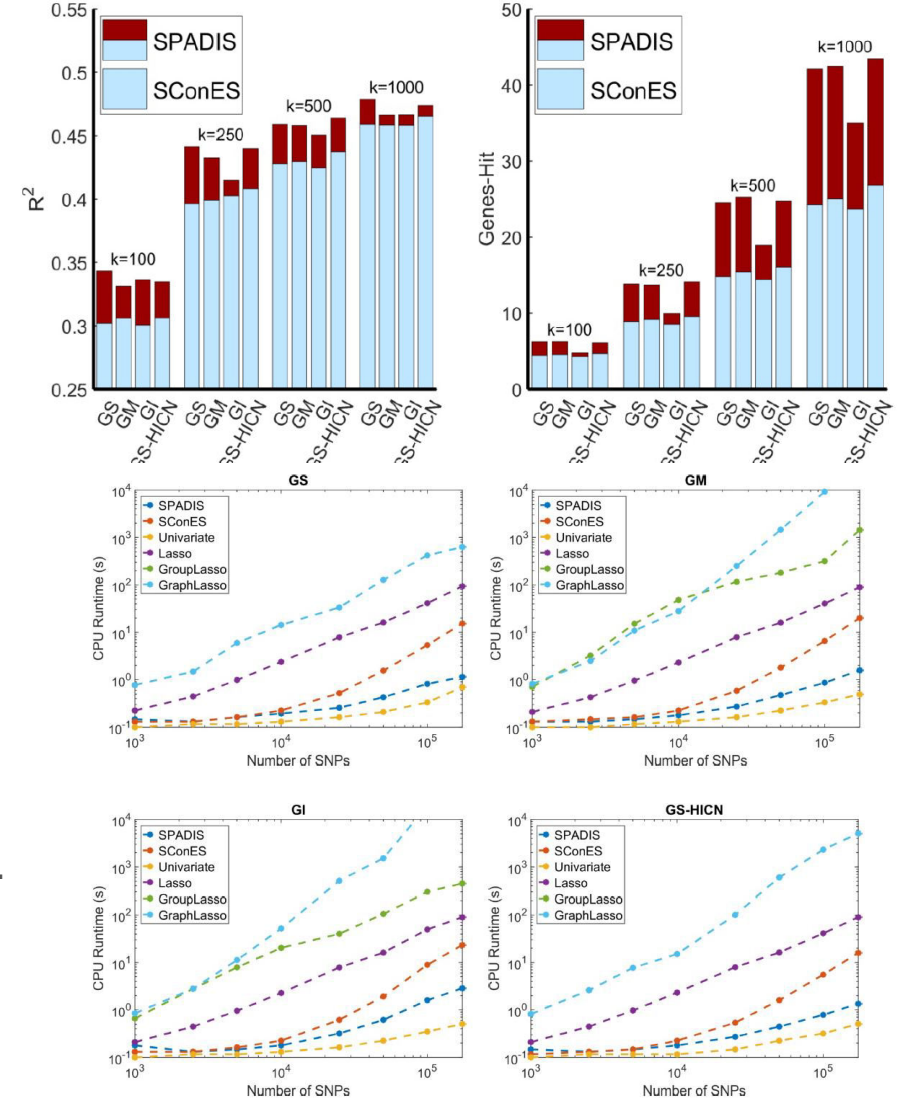
Uzak SNPlerin farklı biyolojik işlevleri etkileme ihtimali daha fazla.

SPADIS



Yakın SNPlar tekrarlayan bilgi getirir.

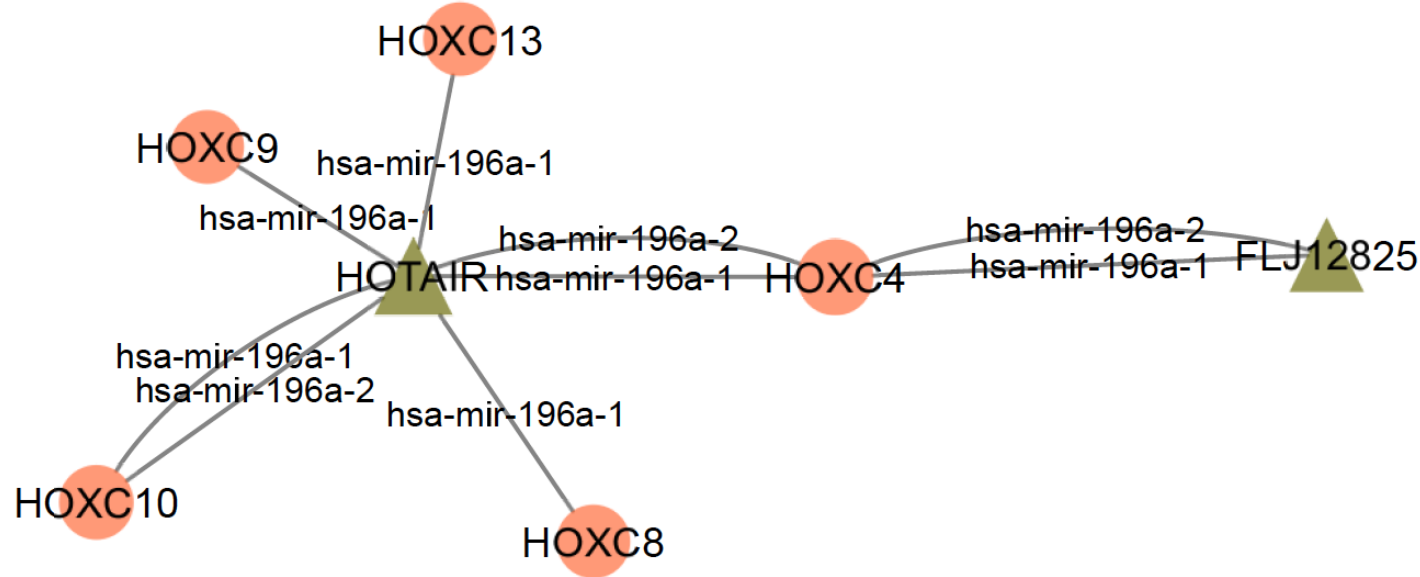
Uzak SNPlerin farklı biyolojik işlevleri etkileme ihtimali daha fazla.



RNA Etkileşimlerinin Çekirdekli Testler ile Tespiti



Gulden Olgun
Bilkent Üniversitesi



Teşekkürler!

TUBİTAK

- 117E140'nolu araştırma projesi

Bilim Akademisi Genç Bilim İnsanları Ödül Programı (2014-2016)

E-mail: otastan@sabanciuniv.edu