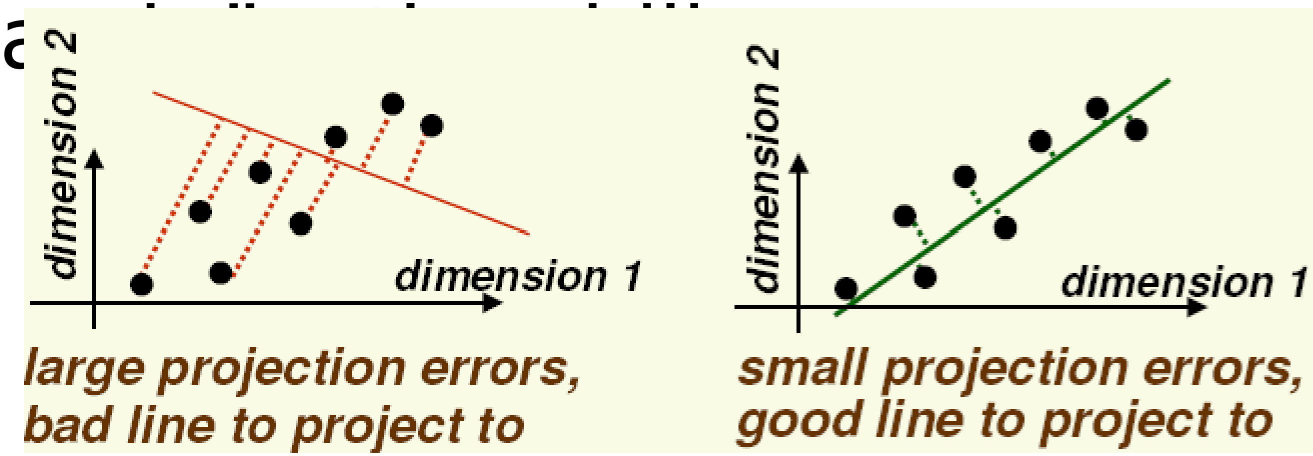
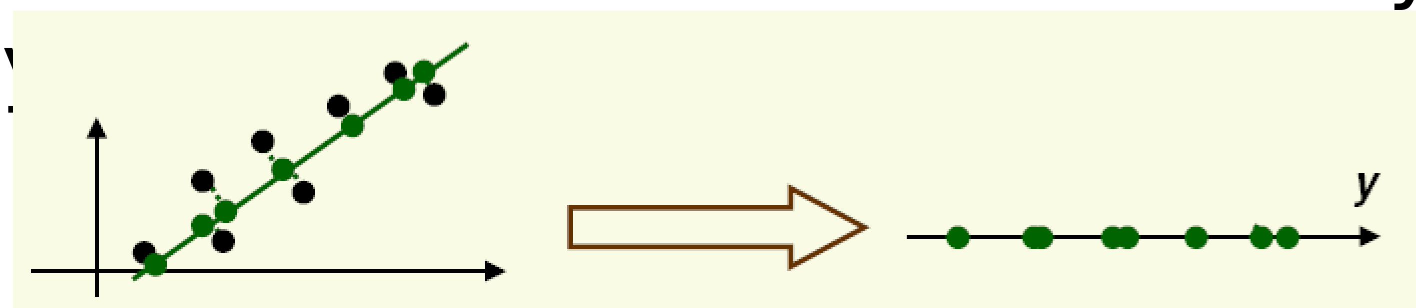
PRINCIPAL COMPONENT ANALYSIS(PCA)

Ana Fikir: Daha az boyutlu uzayda daha doğru veri temsili elde etmek ola



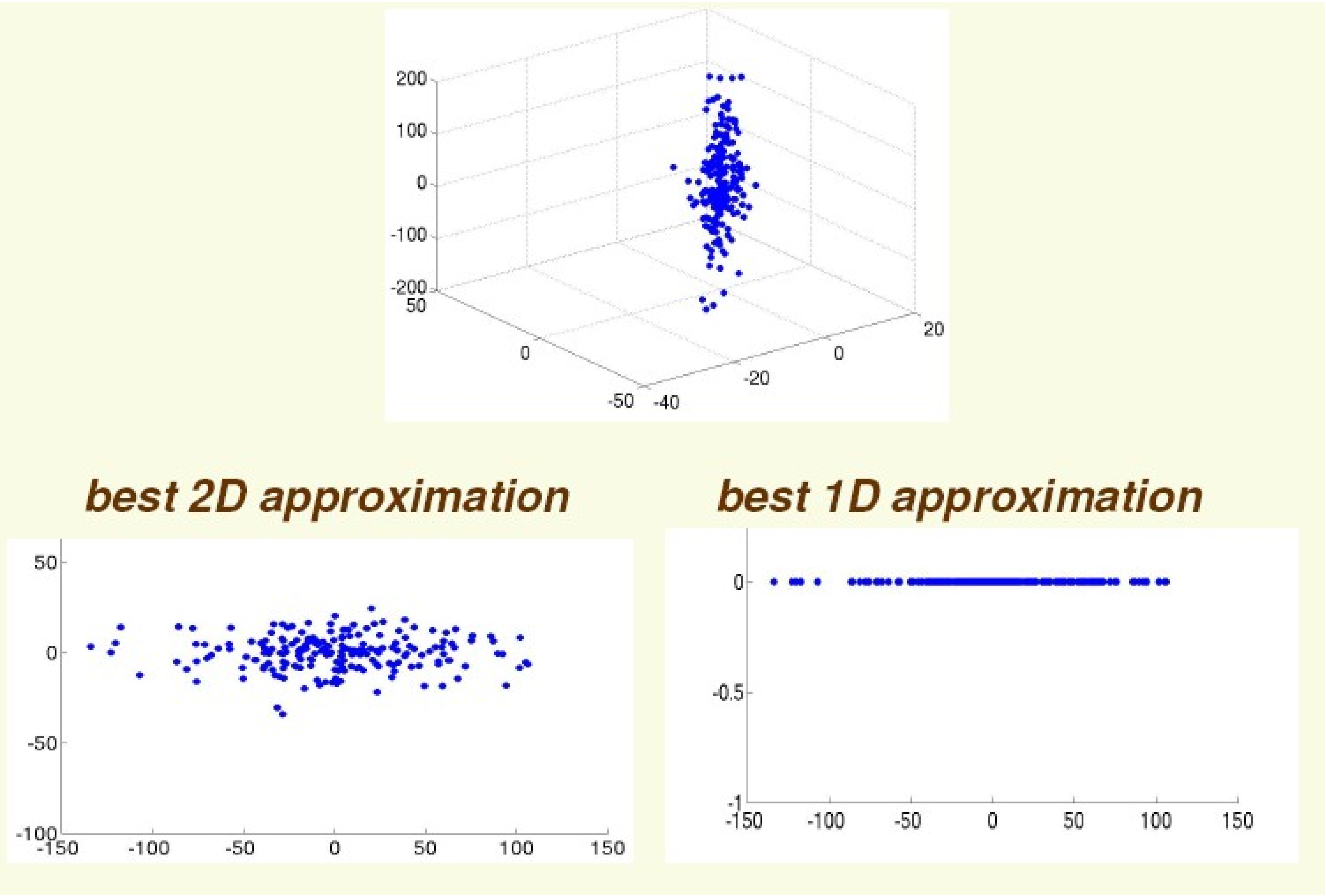
En geniş variance (değişim) aralığında uzanan tek boyutlu bir değişken (çizgi) verilerimizi en doğru temsil edebilen bir yapı olacaktır.

Verilerimizi eniyi temsil edebilecek tek boyutlu yapı tesit edildikten sonra y vektörü ile temsil edilecek bir boyutlu koordinat sistemine transfer edilmeye ihti



y nin yeşil çizgi boyunca uzanan eski **x** verileriyle aynı variance sahip olduğu bilinmektedir.

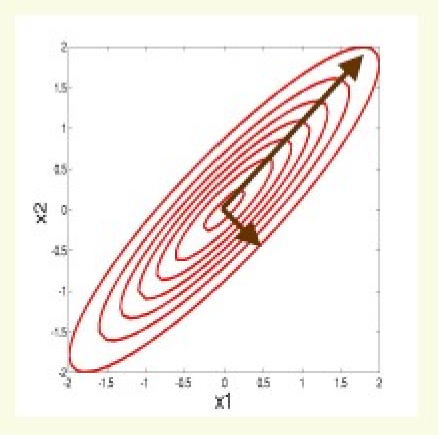
PCA verideki en büyük değişimi korur. Şimdi bu durumu açıklayacağız.



3 boyutta eliptik bulut yaklaşımı

PCA

- Verideki en büyük değişimin yönü nedir?
- Eğer **X** çoklu değişim dağılımı N($\boldsymbol{\mu}, \boldsymbol{\Sigma}$) na sahipse, enbüyük değişimin yönü enbüyük özdeğere ait özvektör tarafından belirlenir.



Bu bizim verinin kovaryans matrisinde bakabileceğimiz nicelik için bir ip ucudur. (PCA Gaussian dağılımın dışındaki dağılımlara dauygulanabilir.)

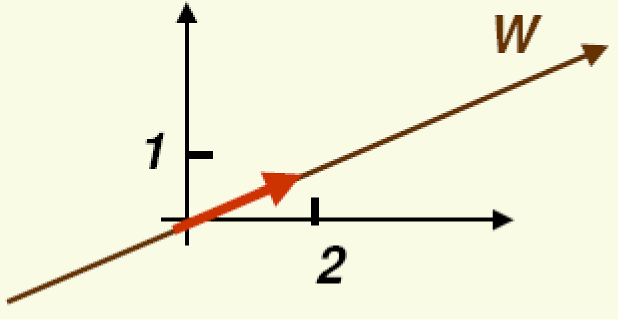
PCA : Türev için Lineer Cebir

- **V** ,**d** boyutlu bir lineer uzay ve **W**, **k** boyutlu ve **V** nin bir altuzayı olsun.
- W için ortonormal bazla $\{e_1, e_2, \ldots, e_k\}$ oluşan d boyutlu vektör kümesi her zaman bulabiliriz.

$\langle e_i, e_j \rangle = 0$ eğer i; $\langle e_i, e_i \rangle = 1$

Böylece **W** alt uzayındaki herhangi bir vektör aşağıdaki gibi yazılabilir

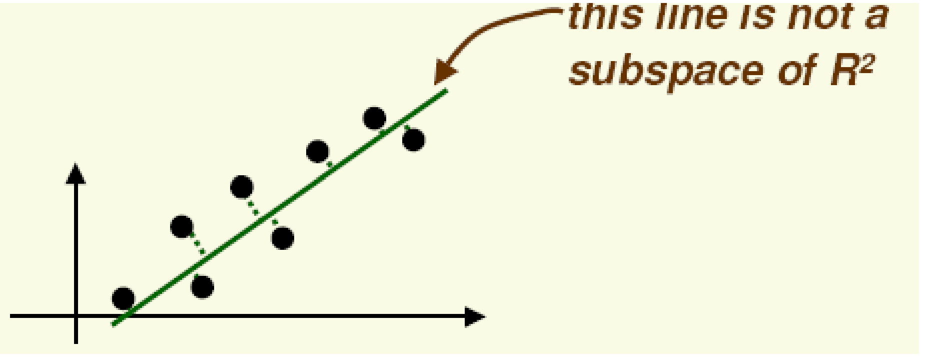
$$\alpha_1 e_1 + \alpha_2 e_2 + \ldots + \alpha_k e_k = \sum_{i=1} \alpha_i e_i \quad \textit{for scalars } \alpha_1, \ldots, \alpha_k$$



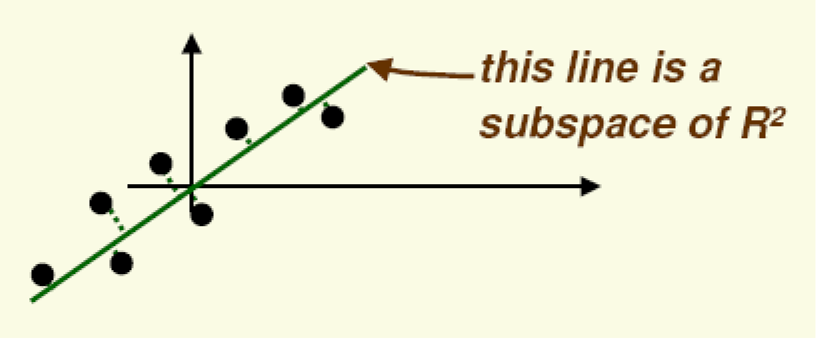
Let $V = R^2$ and W be the line x-2y=0. Then the orthonormal basis for W is

$$\left\{ \begin{bmatrix} 1/\sqrt{5} \\ -2/\sqrt{5} \end{bmatrix} \right\}$$

W uzayının sıfır vektörünü içerdiğini farzedelim. bu durumda orijinden geçecektir.



Türev için, **W** alt uzayına geçmek uygun olacaktır; herşeyi kaydırmamız gerekecektir.



PCA Türev: Ortalama Vektör Kadar Kaydırma

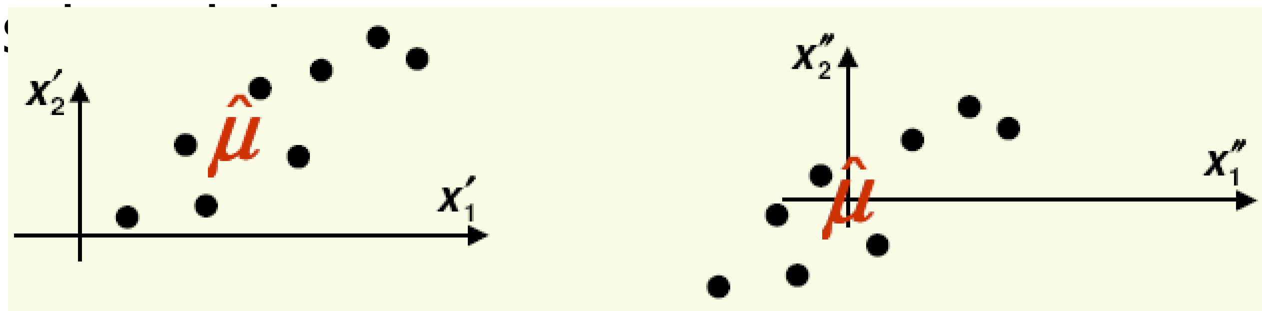
PCA dan önce veriden ortalama örneğini çıkart

$$x - \frac{1}{n}\sum_{i=1}^{n} x_i = x - \hat{\mu}$$

Yeni veriler sıfır ortalamaya sah

$$E(\boldsymbol{X}\text{-}E(\boldsymbol{X})) = E(\boldsymbol{X})\text{-}E(\boldsymbol{X}) = 0$$

Burada bütün yaptığımız koordinat sistemini değiş



Başka bir açıdan **Y** yi elde etmenin ilk adımı **X** in ortalamasını çıkartmak olmuştur.

$$x \rightarrow y = f(x) = g(x - \hat{\mu})$$

PCA Türev

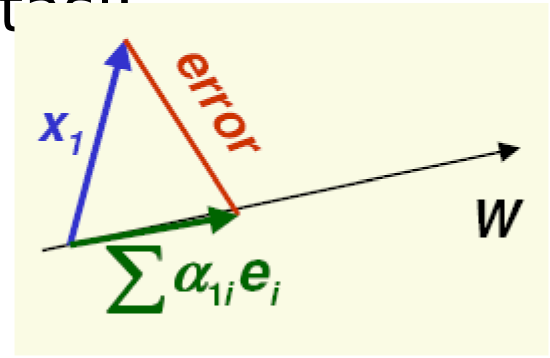
Biz k<d boyutlu bir W altuzayında verimizin $D=\{x_1,x_2,\dots,x_n\}$ temsilini arıyoruz

- $\{e_1,e_2,\dots,e_k\}$ W altuzayının ortonormal bazları olsun. W altuzayındaki herhan vektör $\sum_{i=1}^{k}\alpha_i e_i$ şeklinde yazılabilir.

- Böylece X W uz $\sum_{i=1}^{k}\alpha_{1i} e_i$ şeklinde temsil edilebilir.

$$error=\left\| x_1-\sum_{i=1}^{k}\alpha_{1i} e_i \right\|^2$$

- Bu temsilin hat şeklinde oluşur.



- Toplam hatayı bulmak için bütün xj ler üzerinden toplam alıyoruz.

- Herhangi bir xj $\sum_{i=1}^{k} \alpha_{ji} e_i$ şeklinde yazılabilir.

- Böylece bütün D verilerinin toplam hatası:

$$\underbrace{J(e_1, \ldots, e_k, \alpha_{11}, \ldots \alpha_{nk})}_{\text{unknowns}} = \overbrace{\sum_{j=1}^{n}}^{\text{sum over all data points}} \underbrace{\left\| x_j - \sum_{i=1}^{k} \alpha_{ji} e_i \right\|^2}_{\text{error at one point}}$$

- J yi minimize etmek $\{e_1, e_2, \ldots, e_k\}$ baz vektörlerinin de ortogonal olması gerekliliğinide gözönünde

$$J(e_1,\ldots,e_k,\alpha_{11},\ldots\alpha_{nk}) = \sum_{j=1}^{n} \left\| x_j - \sum_{i=1}^{k} \alpha_{ji} e_i \right\|^2$$

- Öncelikle J nin açılımını yaparsak

$$J(e_1,\ldots,e_k,\alpha_{11},\ldots\alpha_{nk}) = \sum_{j=1}^{n} \|x_j\|^2 - 2\sum_{j=1}^{n} x_j^t \left(\sum_{i=1}^{k} \alpha_{ji} e_i \right) + \sum_{j=1}^{n} \sum_{i=1}^{k} \alpha_{ji}^2$$

$$= \sum_{j=1}^{n} \|x_j\|^2 - 2\sum_{j=1}^{n} \sum_{i=1}^{k} \alpha_{ji} x_j^t e_i + \sum_{j=1}^{n} \sum_{i=1}^{k} \alpha_{ji}^2$$

$$J(e_1,\ldots,e_k,\alpha_{11},\ldots\alpha_{nk}) = \sum_{j=1}^{n} \|x_j\|^2 - 2\sum_{j=1}^{n} \sum_{i=1}^{k} \alpha_{ji} x_j^t e_i + \sum_{j=1}^{n} \sum_{i=1}^{k} \alpha_{ji}^2$$

α_{ml} e göre kısmi türev a $\frac{\partial}{\partial \alpha_{ml}} J(e_1,\ldots,e_k,\alpha_{11},\ldots\alpha_{nk}) = -2x_m^t e_l + 2\alpha_{ml}$

- Böyle α_{ml} için optimal değeri şu şekilde elde ederiz.

$$-2x_m^t e_l + 2\alpha_{ml} = 0 \;\Rightarrow\; \alpha_{ml} = x_m^t e_l$$

$$J(e_1,\ldots,e_k,\alpha_{11},\ldots\alpha_{nk}) = \sum_{j=1}^{n}\left\|x_j\right\|^2 - 2\sum_{j=1}^{n}\sum_{i=1}^{k}\alpha_{ji}x_j^t e_i + \sum_{j=1}^{n}\sum_{i=1}^{k}\alpha_{ji}^2$$

- $\alpha_{ml} = x^t{}_m e_l$ J de yerine yazılırsa

$$J(e_1,\ldots,e_k) = \sum_{j=1}^{n}\left\|x_j\right\|^2 - 2\sum_{j=1}^{n}\sum_{i=1}^{k}\left(x_j^t e_i\right)x_j^t e_i + \sum_{j=1}^{n}\sum_{i=1}^{k}\left(x_j^t e_i\right)^2$$

$$J(e_1,\ldots,e_k) = \sum_{j=1}^{n}\left\|x_j\right\|^2 - \sum_{j=1}^{n}\sum_{i=1}^{k}\left(x_j^t e_i\right)^2$$

$$J(e_1,\ldots,e_k) = \sum_{j=1}^{n}\left\|x_j\right\|^2 - \sum_{j=1}^{n}\sum_{i=1}^{k}\left(x_j^t e_i\right)^2$$

- $(a^t b)^2 = (a^t b)(a^t b) = (b^t a)(a^t b) = b^t (a a^t) b$

kullanılarak

$$J(e_1,\ldots,e_k) = \sum_{j=1}^{n} \left\| x_j \right\|^2 - \sum_{i=1}^{k} e_i^t \left(\sum_{j=1}^{n} \left(x_j x_j^t \right) \right) e_i$$

$$= \sum_{j=1}^{n} \left\| x_j \right\|^2 - \sum_{i=1}^{k} e_i^t S e_i$$

- Burada $S = \sum_{j=1}^{n} x_j x_j^t$

- S matrisine saçılma matrisi denilmektedir ve sadece kovaryans matrisinin n- 1 çarpılmışıdır.

$$\hat{\Sigma} = \frac{1}{n-1} \sum_{j=1}^{n} \left(x_j - \hat{\mu} \right) \left(x_j - \hat{\mu} \right)^t$$

$$J(e_1,\ldots,e_k) = \underbrace{\sum_{j=1}^{n} \|x_j\|^2}_{constant} - \sum_{i=1}^{k} e_i^t S e_i$$

- J yi minimize etm $\sum_{i=1}^{k} e_i^t S e_i$ değerini maksimize etmekle eşdeğerdir.
- Bütün İ ler içi $e_i^t e_i = 1$ şartını koruyarak.
- Lagrange çarpanları yönteminide kullanarak
- yeni bir U fonksiyonunu maksimize etmeliyiz.

$$u(e_1,\ldots,e_k) = \sum_{i=1}^{k} e_i^t S e_i - \sum_{j=1}^{k} \lambda_j \left(e_j^t e_j - 1\right)$$

- Eğer X bir vektör $f(x) = f(x_1, \ldots, x_d)$ bir fonksiyon ise

$$\frac{d}{dx} f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_d} \end{bmatrix}$$

olacağı aşikardır

$$\frac{d}{dx}\left(x^t x\right) = 2x$$

bilgisini kullanarak

- Eğer A bir simetrik matris ise $\frac{d}{dx}\left(x^t A x\right) = 2Ax$ şeklinde kullanılabileceği bilgisine ulaşırız.

$$u(e_1,\ldots,e_k)=\sum_{i=1}^{k} e_i^t S e_i - \sum_{j=1}^{k} \lambda_j \left(e_j^t e_j - 1\right)$$

- e_m ye göre kısmı türevini alacak olursak.

$$\frac{\partial}{\partial e_m} u(e_1,\ldots,e_k) = 2Se_m - 2\lambda_m e_m = 0$$

- Böylece λ_m ve e_m S matrisinin özdeğer ve özvektörleri olmaktadır.

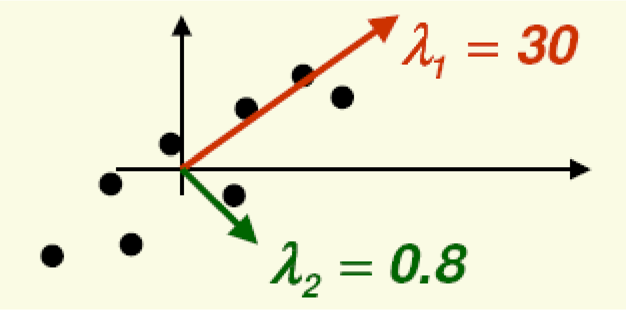
$$Se_m = \lambda_m e_m$$

$$J(e_1,\dots,e_k) = \sum_{j=1}^{n} \|x_j\|^2 - \sum_{i=1}^{k} e_i^t S e_i$$

- Bunları şimdi tekrar J de yerine yazalım.

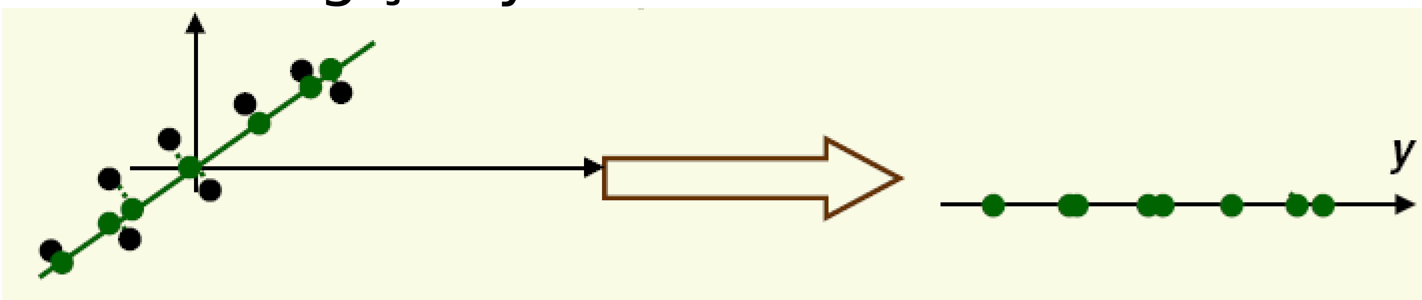
$$J(e_1,\dots,e_k) = \sum_{j=1}^{n} \|x_j\|^2 - \sum_{i=1}^{k} \lambda_i \|e_i\|^2 = \underbrace{\sum_{j=1}^{n} \|x_j\|^2}_{\text{constant}} - \sum_{i=1}^{k} \lambda_i$$

- En büyük özdeğer variance (değişim) ın en büyük olduğu yöndeki özvektörle ilintilidir.



SON ADIM

Nihayi Y vektörünü elde etmek için koordinat sistemini değiştiriyoruz.



- E matrisini oluşturuyo $\boldsymbol{E} = [\boldsymbol{e}_1 \cdots \boldsymbol{e}_k]$

- Koordinat dönüşümünü yapıyo $\boldsymbol{y} = \boldsymbol{E}^t \boldsymbol{x}$
- Özvektörlerimiz yeni koordinat sistemimizin baz vektörleri oluyo

$$\boldsymbol{E}^t \boldsymbol{e}_i = \begin{bmatrix} \boldsymbol{e}_1 \\ \vdots \\ \boldsymbol{e}_i \\ \vdots \\ \boldsymbol{e}_k \end{bmatrix} \boldsymbol{e}_i = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix}$$

Recipe for Dimension Reduction with PCA

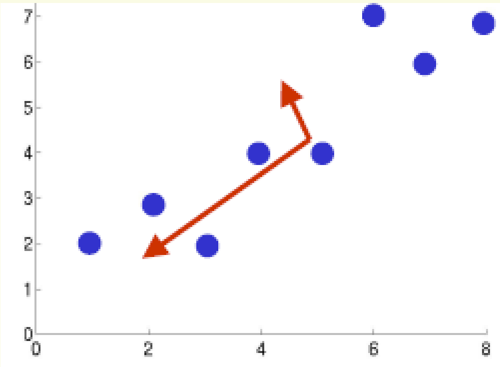
Data $\boldsymbol{D}=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\ldots,\boldsymbol{x}_n\}$. Each $\boldsymbol{x}_i$ is a $\boldsymbol{d}$-dimensional vector. Wish to use PCA to reduce dimension to $\boldsymbol{k}$

1. Find the sample mean $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i$
2. Subtract sample mean from the data $\boldsymbol{z}_i = \boldsymbol{x}_i - \hat{\mu}$
3. Compute the scatter matrix $\boldsymbol{S} = \sum_{i=1}^{n} \boldsymbol{z}_i \boldsymbol{z}_i^t$
4. Compute eigenvectors $\boldsymbol{e}_1,\boldsymbol{e}_2,\ldots,\boldsymbol{e}_k$ corresponding to the $\boldsymbol{k}$ largest eigenvalues of $\boldsymbol{S}$
5. Let $\boldsymbol{e}_1,\boldsymbol{e}_2,\ldots,\boldsymbol{e}_k$ be the columns of matrix $\boldsymbol{E} = [\boldsymbol{e}_1 \cdots \boldsymbol{e}_k]$
6. The desired $\boldsymbol{y}$ which is the closest approximation to $\boldsymbol{x}$ is $\boldsymbol{y} = \boldsymbol{E}^t \boldsymbol{z}$

PCA Example Using Matlab

- Let $\boldsymbol{D} = \{(1,2),(2,3),(3,2),(4,4),(5,4),(6,7),(7,6),(9,7)\}$
- Convenient to arrange data in array

$$\boldsymbol{X} = \begin{bmatrix} 1 & 2 \\ \vdots & \vdots \\ 9 & 7 \end{bmatrix} = \begin{bmatrix} x_1 \\ \vdots \\ x_8 \end{bmatrix}$$



- Mean $\mu = \boldsymbol{mean}(\boldsymbol{X}) = [4.6 \quad 4.4]$
- Subtract mean from data to get new data array Z

$$\boldsymbol{Z} = \boldsymbol{X} - \begin{bmatrix} \mu \\ \vdots \\ \mu \end{bmatrix} = \boldsymbol{X} - \boldsymbol{repmat}(\mu, 8, 1) = \begin{bmatrix} -3.6 & -4.4 \\ \vdots & \vdots \\ 4.4 & 2.6 \end{bmatrix}$$

- Compute the scatter matrix $\boldsymbol{S}$

$$\boldsymbol{S} = 7 * \boldsymbol{cov}(\boldsymbol{Z}) = [-3.6 \quad -4.4]\begin{bmatrix} -3.6 \\ -4.4 \end{bmatrix} + \ldots + [4.4 \quad 2.6]\begin{bmatrix} 4.4 \\ 2.6 \end{bmatrix} = \begin{bmatrix} 57 & 40 \\ 40 & 34 \end{bmatrix}$$

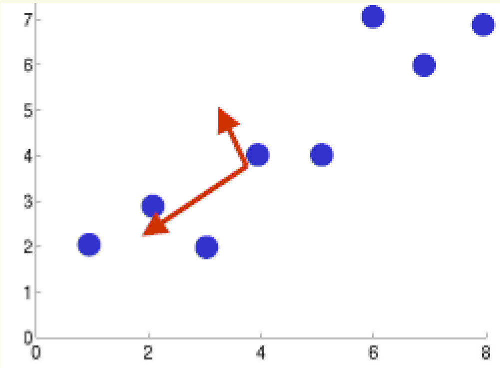
*matlab uses unbiased estimate for covariance, so S=(n-1)*cov(Z)*

PCA Example Using Matlab

- Use [V,D] =eig($\boldsymbol{S}$) to get eigenvalues and eigenvectors of $\boldsymbol{S}$

$$\lambda_1 = 87 \text{ and } e_1 = \begin{bmatrix} -0.8 \\ -0.6 \end{bmatrix}$$

$$\lambda_2 = 3.8 \text{ and } e_2 = \begin{bmatrix} 0.6 \\ -0.8 \end{bmatrix}$$



- Projection to 1D space in the direction of $\boldsymbol{e_1}$

$$Y = e_1^t Z^t = \left(\begin{bmatrix} -0.8 & -0.6 \end{bmatrix} \begin{bmatrix} -3.6 & \cdots & 4.4 \\ -4.4 & \cdots & 2.6 \end{bmatrix} \right) = \begin{bmatrix} 4.3 & \cdots & -5.1 \end{bmatrix}$$

$$= \begin{bmatrix} y_1 & \cdots & y_8 \end{bmatrix}$$