

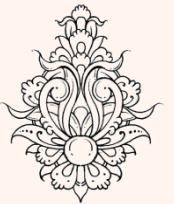
یادگیری ماشین  
(۰۱-۸۰۵-۱۱-۱۳)  
فصل پنجم



دانشگاه شهید بهشتی  
پژوهشکده‌ی فضای مجازی  
پاییز ۱۳۹۶  
احمد محمودی ازناوه

# فهرست مطالب

- داده‌های چندمتغیره
- تخمین پارامترها
- کواریانس و ضریب همبستگی
- دسته‌بندی
- رگرسیون



# داده‌های چندمتغیره

- در بسیاری از کاربردها، اندازه‌گیری‌های متفاوتی انجام می‌شود، از این جهت با بردار ورودی (ویژگی) سروکار خواهیم داشت (به عنوان مثال یک بردار  $d$ -بعدی).

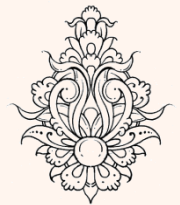
*Data matrix*

$$\mathbf{X} = \begin{bmatrix} X_1^1 & X_2^1 & \dots & X_d^1 \\ X_1^2 & X_2^2 & \dots & X_d^2 \\ \vdots & & & \\ X_1^N & X_2^N & \dots & X_d^N \end{bmatrix}$$

یک نمونه

*d inputs/features/attributes*

*N instances/observations/examples*

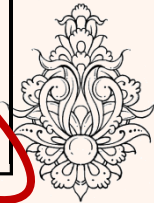


# پارامترهای چندمتغیره

$$\text{Mean: } E[\mathbf{x}] = \boldsymbol{\mu} = [\mu_1, \dots, \mu_d]^T$$

$$\sigma_i \equiv \text{var}(X_i)$$

$$\text{Cov}(X_i, X_j) = E[(X_i - \mu_i)(X_j - \mu_j)] = E[X_i X_j] - \mu_i \mu_j$$

$$\begin{aligned} \Sigma \equiv \text{Cov}(X) &= E[(X - \boldsymbol{\mu})(X - \boldsymbol{\mu})^T] = \\ &= E[XX^T] - \boldsymbol{\mu}\boldsymbol{\mu}^T \end{aligned} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1d} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{d1} & \sigma_{d2} & \dots & \sigma_d^2 \end{bmatrix}$$


$$\mathbf{X} = \begin{bmatrix} X_1^1 & X_2^1 & \dots & X_d^1 \\ X_1^2 & X_2^2 & \dots & X_d^2 \\ \vdots & \vdots & \ddots & \vdots \\ X_1^N & X_2^N & \dots & X_d^N \end{bmatrix}$$

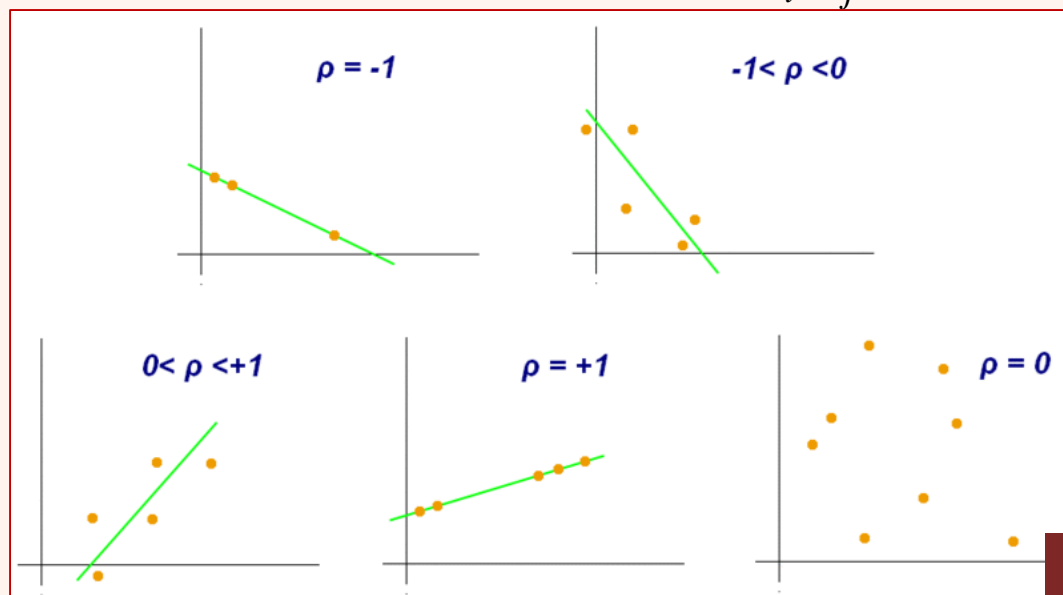
واریانس



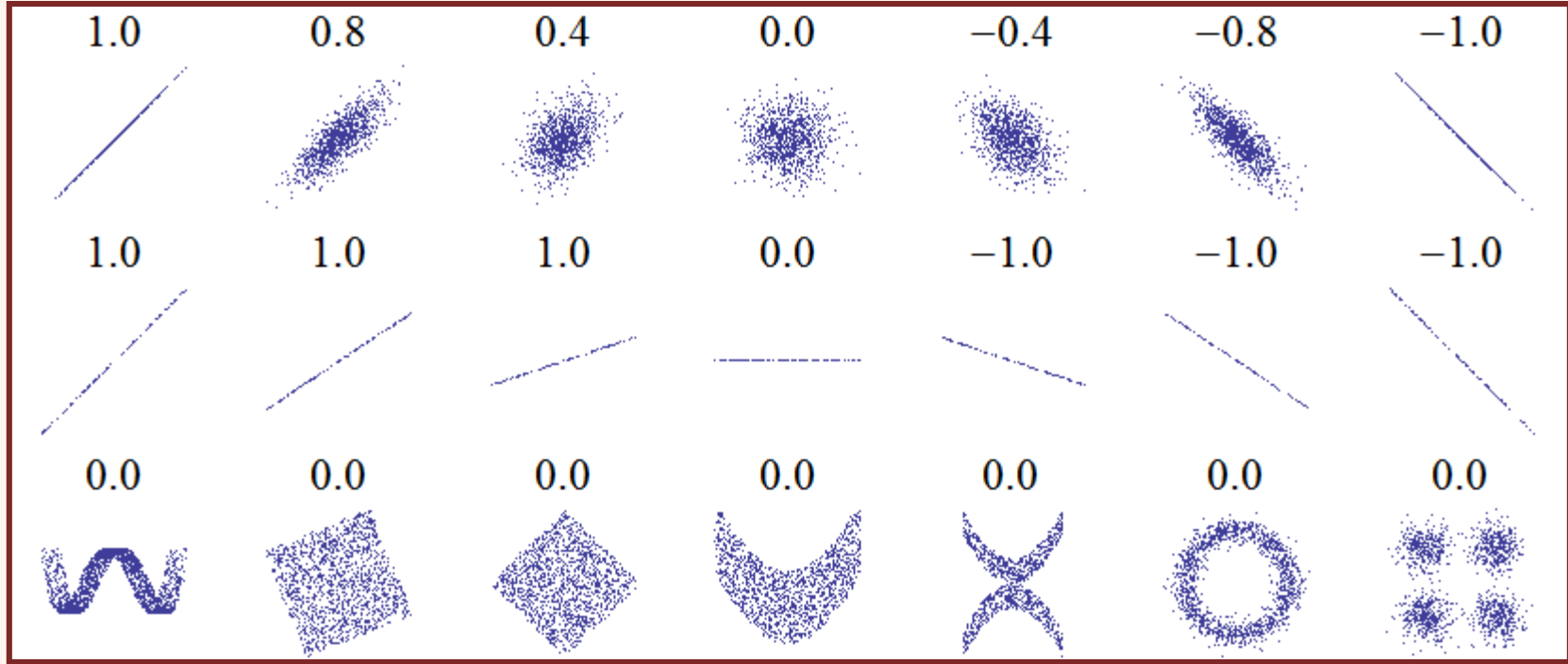
# ضریب همبستگی Pearson

- میزان همبستگی خطی بین دو متغیر تصادفی را می‌سنجد.
  - مقدار این ضریب بین -1 تا 1 تغییر می‌کند که «1» به معنای همبستگی مثبت کامل، «0» به معنی نبود همبستگی، و «-1» به معنی همبستگی منفی کامل است.
  - این ضریب که کاربرد فراوانی در آمار دارد، توسط Karl Pearson براساس ایده‌ی اولیه‌ی Francis Galton تدوین شد.

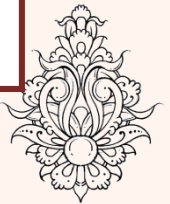
$$\text{Correlation: } \text{Corr}(X_i, X_j) \equiv \rho_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j}$$



# مثال



wikipedia

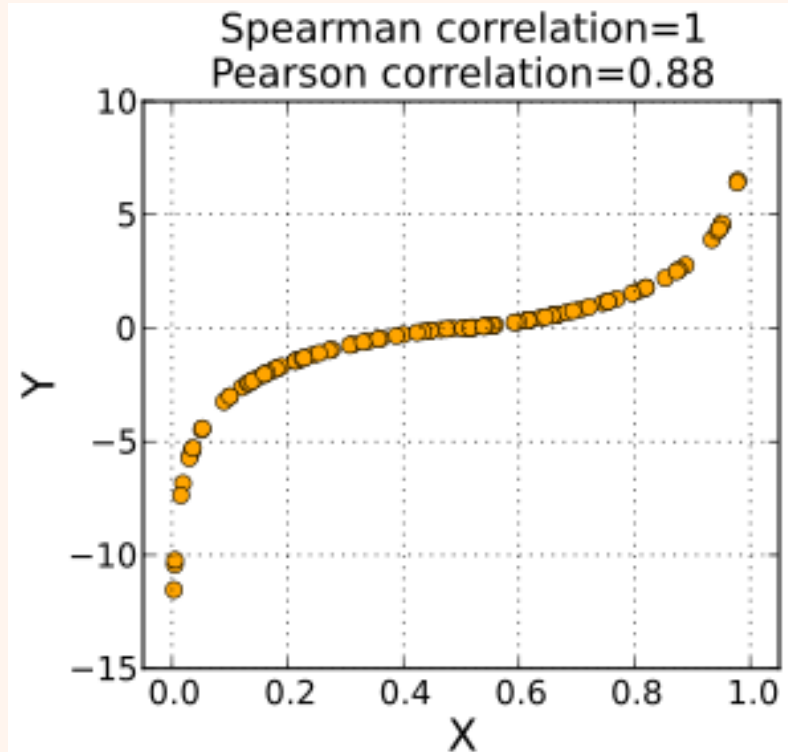


در صورتی که دو متغیر مستقل باشند، کواریانس و در نتیجه ضریب همبستگی آن دو صفر خواهد بود. اما عکس این مسأله درست نیست.

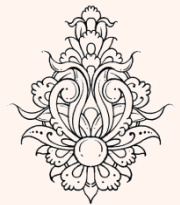


# ضریب همبستگی رتبه‌ای Spearman

## Spearman's rank correlation coefficient



$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}$$

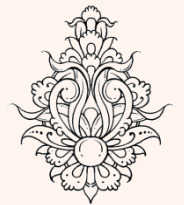


# تخمین پارامترها

Sample mean  $\mathbf{m}$  :  $m_i = \frac{\sum_{t=1}^N x_i^t}{N}, i = 1, \dots, d$

Covariance matrix  $\mathbf{S}$  :  $s_{ij} = \frac{\sum_{t=1}^N (x_i^t - m_i)(x_j^t - m_j)}{N}$

Correlation matrix  $\mathbf{R}$  :  $r_{ij} = \frac{s_{ij}}{s_i s_j}$





# تخمین پارامترهای نامشخص

## Estimation of Missing Values

- ممکن است در برخی نمونه‌ها، برخی متغیرها در اختیار نباشند.

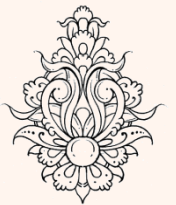
- بهترین راه صرفنظر کردن از آنهاست، اما این راه در حالتی که داده‌های آموزشی محدود باشد، کارایی ندارد.

- یک فیلد جدید اضافه کنیم که فقدان مقدار را مشخص می‌کند؛ ممکن است دارای اطلاعات ارزشمندی باشد.

- نسبت دادن مقدار: (imputation)

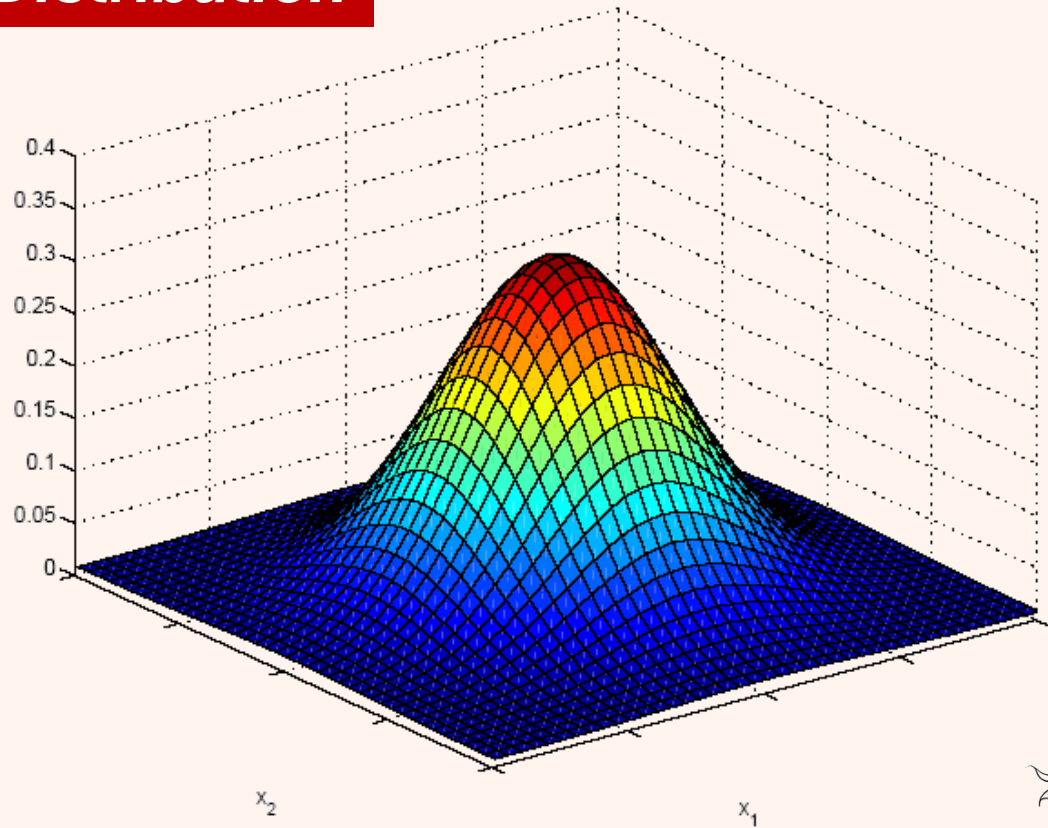
- جایگزینی مقدار میانگین (mean imputation)

- انتساب با رگرسیون (imputation by regression)



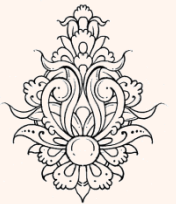
# توزیع نرمال چند متغیره

## Multivariate Normal Distribution



$$\mathbf{x} \sim \mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{d/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right]$$



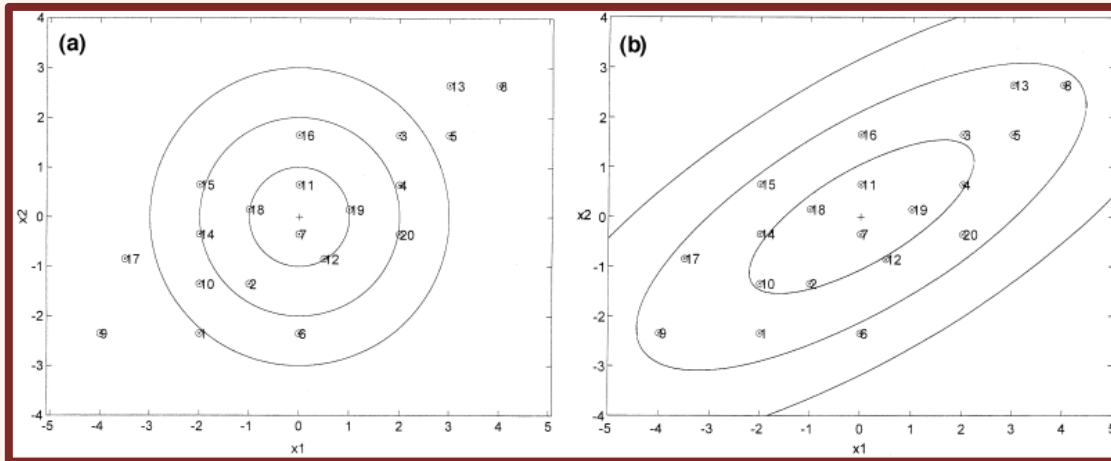
# توزیع نرمال چند متغیره (ادامه...)

Distance in standard units

• فاصله‌ی Mahalanobis:

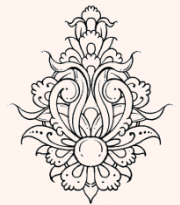
– معیاری برای اندازه‌گیری فاصله‌ی یک نقطه از یک توزیع داده است.

$$(x - \mu)^T \Sigma^{-1} (x - \mu)$$



•  $(x - \mu)^T \Sigma^{-1} (x - \mu) = c^2$

ابریضی



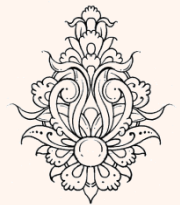
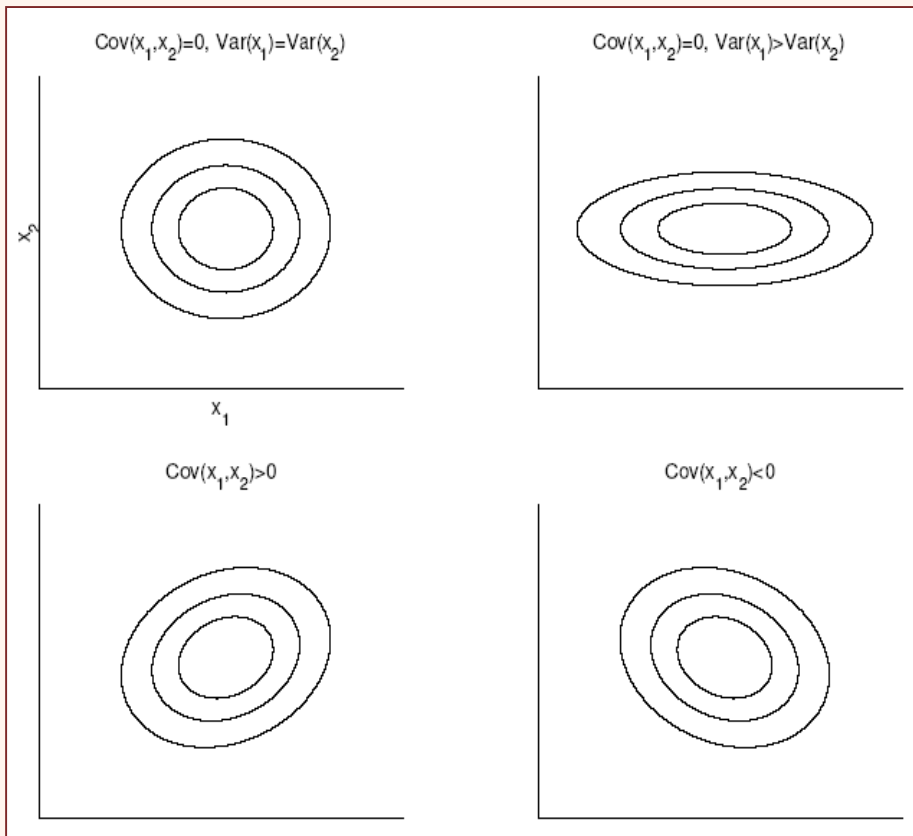
# مثال - دو بعدی

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

$$p(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}(z_1^2 - 2\rho z_1 z_2 + z_2^2)\right]$$

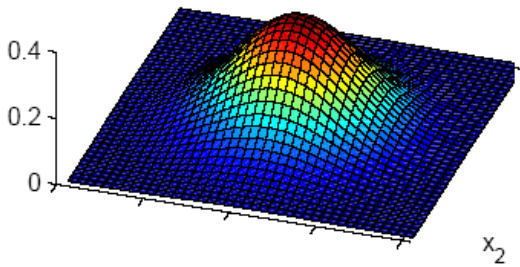
$$z_i = (x_i - \mu_i) / \sigma_i$$

**z normalization**

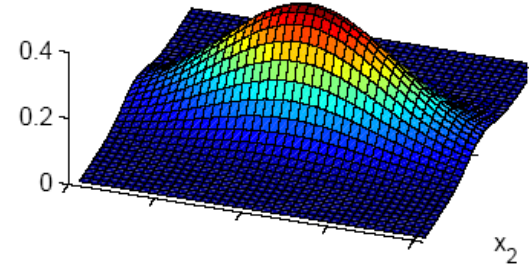


Cov(x<sub>1</sub>,x<sub>2</sub>)=0, Var(x<sub>1</sub>)=Var(x<sub>2</sub>)

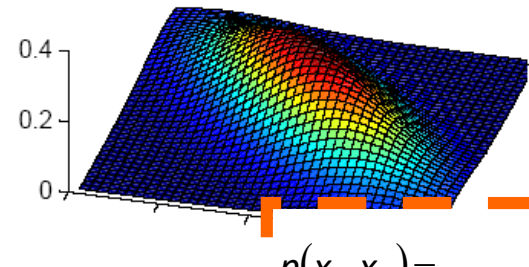
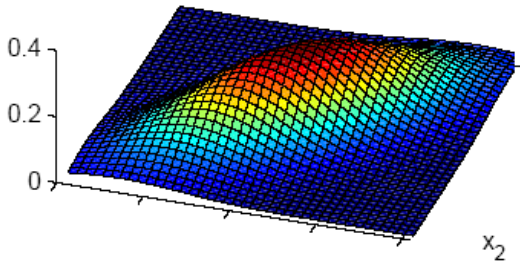
Cov(x<sub>1</sub>,x<sub>2</sub>)=0, Var(x<sub>1</sub>)>Var(x<sub>2</sub>)



Cov(x<sub>1</sub>,x<sub>2</sub>)>0



Cov(x<sub>1</sub>,x<sub>2</sub>)<0



$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

$$p(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}(z_1^2 - 2\rho z_1 z_2 + z_2^2)\right]$$

$$z_i = (x_i - \mu_i) / \sigma_i$$

**nonsingular**

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})\right]$$

**positive definite**

(var≠0, |ρ|<1)

If not -> Dimension reduction



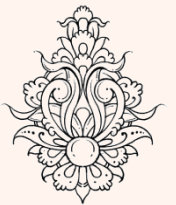
# چند نکته

- در صورتی که  $\mathbf{x}$  دارای توزیع نرمال (چندمتغیره) باشد، متغیر مربوط به هر بعد نیز دارای توزیع نرمال تک‌متغیره است. (عکس این مطلب درست نیست)

– در واقع، هر زیر مجموعه  $k$  بعدی ( $k < d$ ) نیز یک توزیع نرمال چندمتغیره است.

- در صورتی که متغیرها مستقل باشند:

$$p(\mathbf{x}) = \prod_{i=1}^d p_i(x_i) = \frac{1}{(2\pi)^{d/2} \prod_{i=1}^d \sigma_i} \exp \left[ -\frac{1}{2} \sum_{i=1}^d \left( \frac{x_i - \mu_i}{\sigma_i} \right)^2 \right]$$



ادامه ...

$$\mathbf{x} \sim \mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad \mathbf{w} \in \mathbb{R}^d$$

• در صورتی که

$$\mathbf{w}^T \mathbf{x} = w_1 x_1 + w_2 x_2 + \dots + w_d x_d \sim \mathcal{N}(\mathbf{w}^T \boldsymbol{\mu}, \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w})$$

$$E[\mathbf{w}^T \mathbf{x}] = \mathbf{w}^T E[\mathbf{x}] = \mathbf{w}^T \boldsymbol{\mu}$$

$$\begin{aligned} \text{Var}(\mathbf{w}^T \mathbf{x}) &= E\left[(\mathbf{w}^T \mathbf{x} - \mathbf{w}^T \boldsymbol{\mu})^2\right] = E\left[(\mathbf{w}^T \mathbf{x} - \mathbf{w}^T \boldsymbol{\mu})(\mathbf{w}^T \mathbf{x} - \mathbf{w}^T \boldsymbol{\mu})\right] \\ &= E\left[\mathbf{w}^T (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{w}\right] = \mathbf{w}^T E\left[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T\right] \mathbf{w} \\ &= \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w} \end{aligned}$$

در صورتی که  $k$  بردار در نظر گرفته شود ( $k < d$ ):

$\mathbf{w}$  is  $d \times k$

$$\mathbf{w}^T \mathbf{x} \sim \mathcal{N}_k(\mathbf{w}^T \boldsymbol{\mu}, \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w})$$



# دسته بندی چندمتغیره

$$p(\mathbf{x} | C_i) \sim N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$$

$$p(\mathbf{x} | C_i) = \frac{1}{(2\pi)^{d/2} |\boldsymbol{\Sigma}_i|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{x} - \boldsymbol{\mu}_i)\right]$$

## Discriminant functions

$$g_i(\mathbf{x}) = \log p(\mathbf{x} | C_i) + \log P(C_i)$$

$$= -\frac{d}{2} \log 2\pi - \frac{1}{2} \log |\boldsymbol{\Sigma}_i| - \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{x} - \boldsymbol{\mu}_i) + \log P(C_i)$$





# تخمین پارامترها

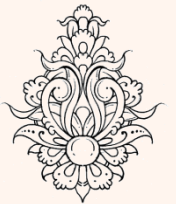
با تخمین و جایگزینی پارامترها خواهیم داشت:

$$\hat{P}(C_i) = \frac{\sum_t r_i^t}{N}$$

$$\mathbf{m}_i = \frac{\sum_t r_i^t \mathbf{x}^t}{\sum_t r_i^t}$$

$$\mathbf{S}_i = \frac{\sum_t r_i^t (\mathbf{x}^t - \mathbf{m}_i)(\mathbf{x}^t - \mathbf{m}_i)^T}{\sum_t r_i^t}$$

$$g_i(\mathbf{x}) = -\frac{1}{2} \log |\mathbf{S}_i| - \frac{1}{2} (\mathbf{x} - \mathbf{m}_i)^T \mathbf{S}_i^{-1} (\mathbf{x} - \mathbf{m}_i) + \log \hat{P}(C_i)$$



# Quadratic discriminant

$$g_i(\mathbf{x}) = -\frac{1}{2} \log |\mathbf{S}_i| - \frac{1}{2} \left( \mathbf{x}^T \mathbf{S}_i^{-1} \mathbf{x} - 2 \mathbf{x}^T \mathbf{S}_i^{-1} \mathbf{m}_i + \mathbf{m}_i^T \mathbf{S}_i^{-1} \mathbf{m}_i \right) + \log \hat{P}(C_i)$$
$$= \mathbf{x}^T \mathbf{W}_i \mathbf{x} + \mathbf{w}_i^T \mathbf{x} + w_{i0}$$

where

$$\mathbf{W}_i = -\frac{1}{2} \mathbf{S}_i^{-1}$$

$$\mathbf{w}_i = \mathbf{S}_i^{-1} \mathbf{m}_i$$

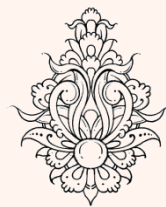
$$w_{i0} = -\frac{1}{2} \mathbf{m}_i^T \mathbf{S}_i^{-1} \mathbf{m}_i - \frac{1}{2} \log |\mathbf{S}_i| + \log \hat{P}(C_i)$$

K.d for means

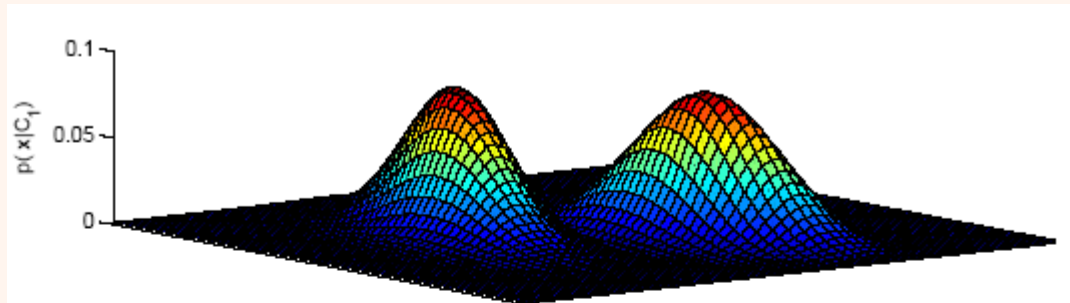
تعداد پارامترها

K.d.(d+1)/2 for covariance

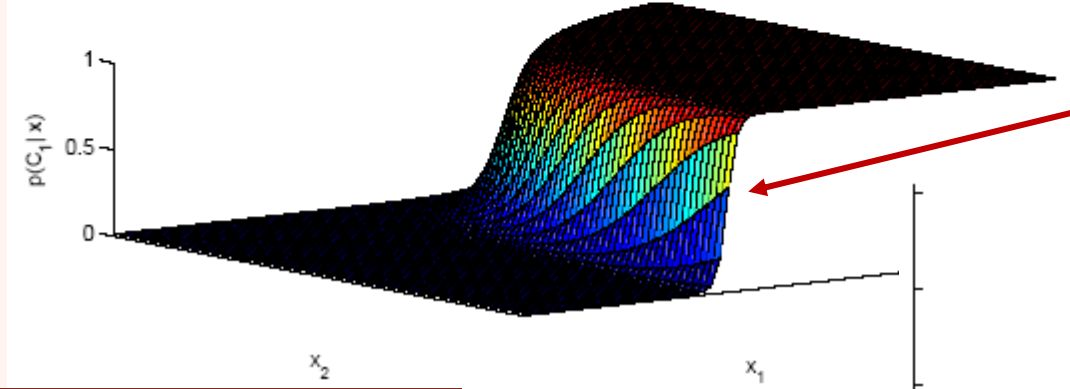
در کم بودن تعداد نمونه‌های آموزشی ممکن است  
ماتریس **singular** شود.



# Quadratic discriminant

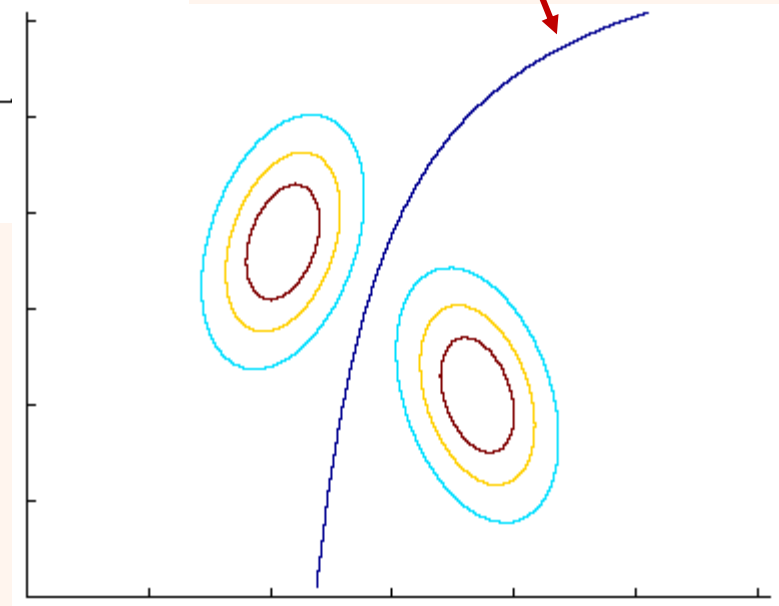


likelihoods



posterior for  $C_1$

discriminant:  
 $P(C_1 | x) = 0.5$



- در صورت کم بودن تعداد نمونه‌های آموزشی می‌توان ماتریس کواریانس را برای همه‌ی

کلاس‌ها یکسان در نظر گرفت.

$$S = \sum_i \hat{P}(C_i) S_i$$

$$g_i(\mathbf{x}) = -\frac{1}{2} \log |S_i| - \frac{1}{2} (\mathbf{x}^T S_i^{-1} \mathbf{x} - 2\mathbf{x}^T S_i^{-1} \mathbf{m}_i + \mathbf{m}_i^T S_i^{-1} \mathbf{m}_i) + \log \hat{P}(C_i)$$

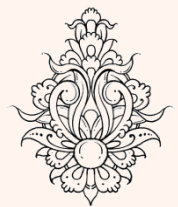
- در نتیجه خواهیم داشت:

$$g_i(\mathbf{x}) = -\frac{1}{2} (\mathbf{x} - \mathbf{m}_i)^T \mathbf{S}^{-1} (\mathbf{x} - \mathbf{m}_i) + \log \hat{P}(C_i)$$

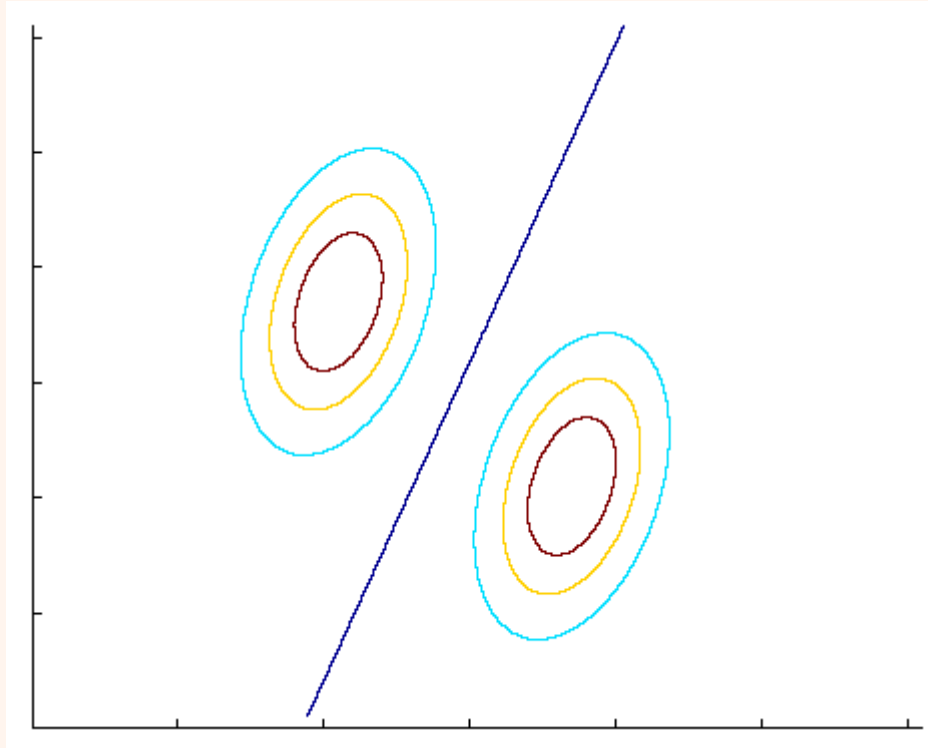
$$g_i(\mathbf{x}) = \mathbf{w}_i^T \mathbf{x} + w_{i0}$$

where

$$\mathbf{w}_i = \mathbf{S}^{-1} \mathbf{m}_i \quad w_{i0} = -\frac{1}{2} \mathbf{m}_i^T \mathbf{S}^{-1} \mathbf{m}_i + \log \hat{P}(C_i)$$



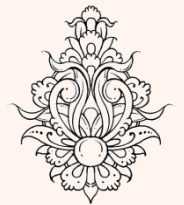
# جداساز خطی



K.d for means

تعداد پارامترها

$d.(d+1)/2$  for covariance



## Diagonal S

• در صورتی که متغیرها، مستقل در نظر گرفته شوند، ماتریس کواریانس قطری خواهد بود:

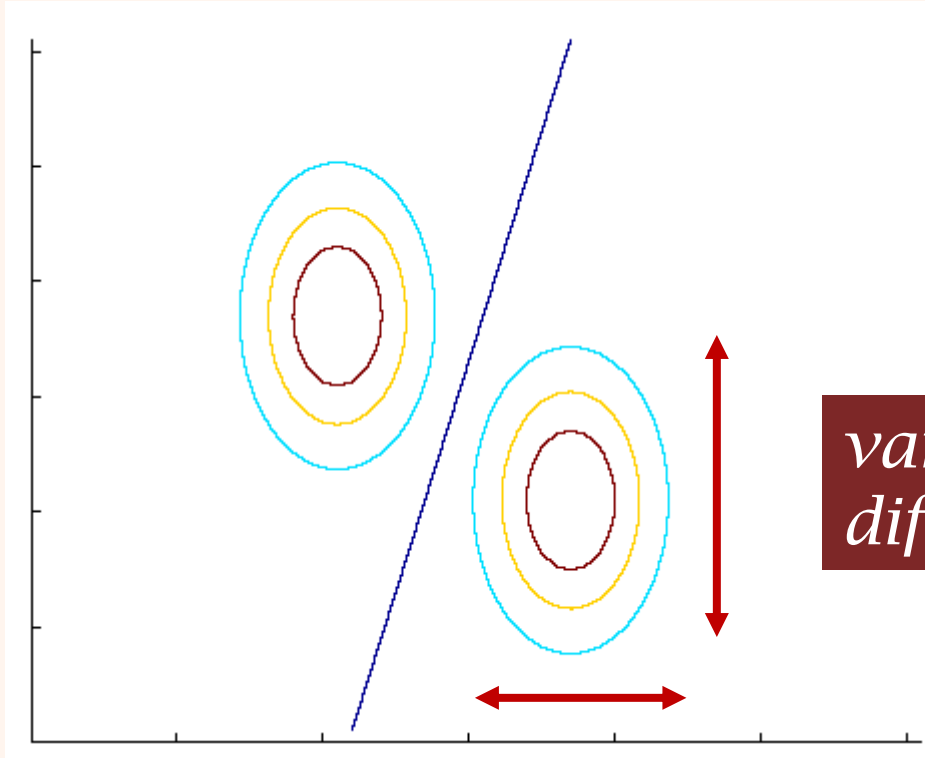
- $p(\mathbf{x} | C_i) = \prod_j p(x_j | C_i)$

## Naïve bayes' classifier

$$g_i(\mathbf{x}) = -\frac{1}{2} \sum_{j=1}^d \left( \frac{x_j^t - m_{ij}}{s_j} \right)^2 + \log \hat{P}(C_i)$$

**weighted Euclidean distance**



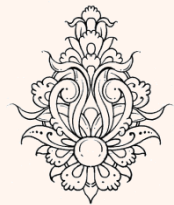


*variances may be different*

k.d for means

d for covariance

تعداد پارامترها



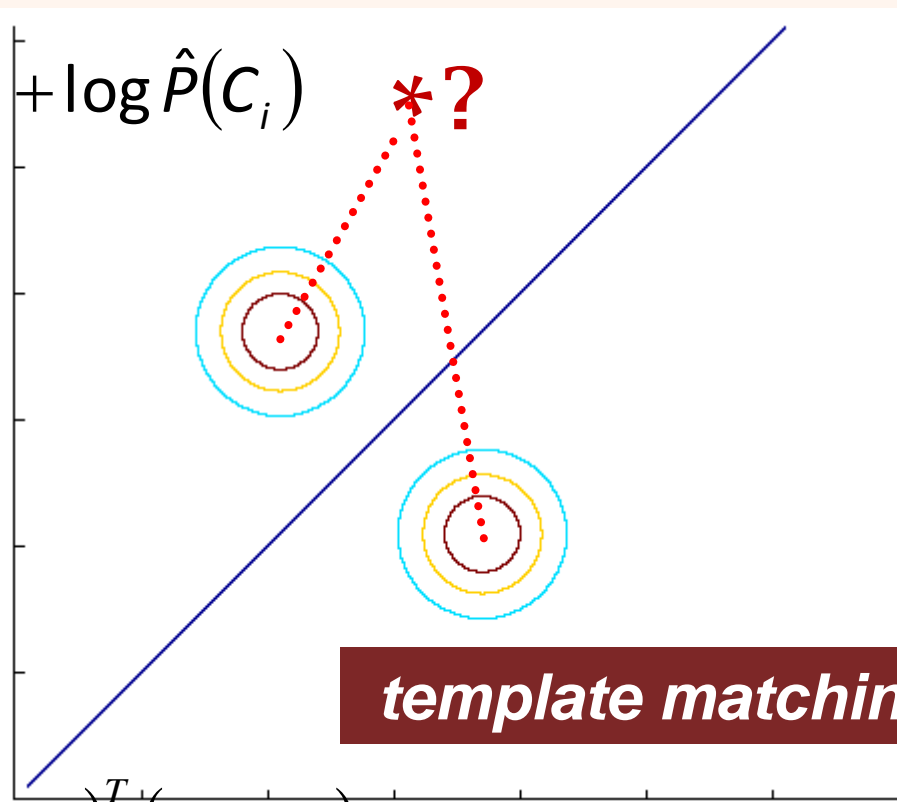
# Nearest mean classifier

در صورتی که واریانس متغیرها هم یکسان در نظر گرفته شود:

$$g_i(\mathbf{x}) = -\frac{\|\mathbf{x} - \mathbf{m}_i\|^2}{2s^2} + \log \hat{P}(C_i)$$

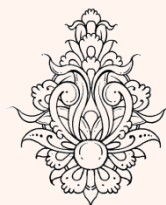
$$= -\frac{1}{2s^2} \sum_{j=1}^d (x_j^t - m_{ij})^2 + \log \hat{P}(C_i)$$

در صورتی که احتمال  
تعلق به کلاسها  
هم اندازه باشند، از  
ضرب داخلی نیز  
می‌توان استفاده کرد



$$g_i(\mathbf{x}) = -\|\mathbf{x} - \mathbf{m}_i\|^2 = -(\mathbf{x} - \mathbf{m}_i)^T (\mathbf{x} - \mathbf{m}_i)$$

$$= -(\mathbf{x}^T \mathbf{x} - 2\mathbf{m}_i^T \mathbf{x} + \mathbf{m}_i^T \mathbf{m}_i)$$





# Tuning Complexity

Assumption	Covariance matrix	No of parameters
Shared, Hyperspheric	$\mathbf{S}_i = \mathbf{S} = s^2 \mathbf{I}$	1
Shared, Axis-aligned	$\mathbf{S}_i = \mathbf{S}$ , with $s_{ij} = 0$	$d$
Shared, Hyperellipsoidal	$\mathbf{S}_i = \mathbf{S}$	$d(d+1)/2$
Different, Hyperellipsoidal	$\mathbf{S}_i$	$K d(d+1)/2$

در نظر گرفتن ماتریس کواریانس مشترک معادل داشتن جداساز خطی است.

فاصله‌ی اقلیدسی زمانی مورد استفاده قرار می‌گیرد، که واریانس هم‌ی متخیرها یکسان در نظر گرفته شود.

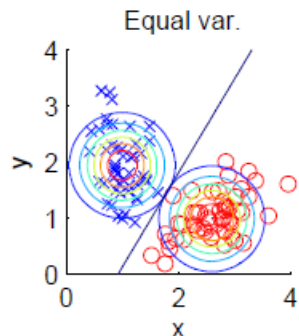
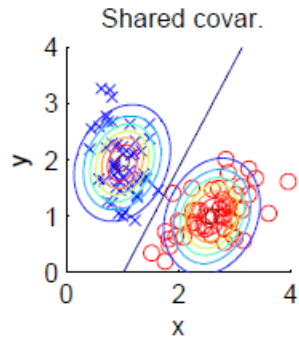
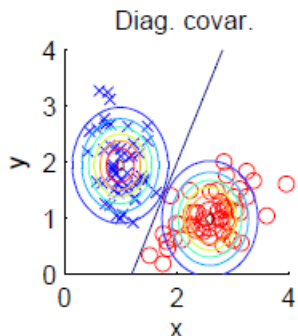
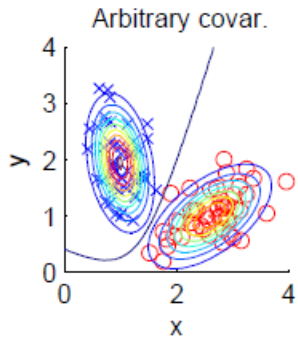
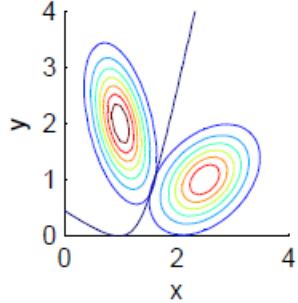


# Regularized discriminant analysis

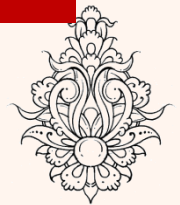
Friedman, J. H. 1989. "Regularized Discriminant Analysis."  
Journal of American Statistical Association 84: 165–175.

$$S'_i = \alpha\sigma^2 I + \beta S + (1 - \alpha - \beta)S_i$$

Population likelihoods and posteriors



با انتخاب مناسب  
 $\alpha$  و  $\beta$  می توان  
پیچیدگی مدل را  
تنظیم کرد.



- در برخی کاربردها، خصیصه‌ها مقداری گسسته دارند،

به عنوان مثال:  $\text{color} \in \{\text{red, blue, green, black}\}$

$\text{pixel} \in \{\text{on, off}\}$

- در صورتی که مقدار اختصاص داده شده دودویی

باشد (توزیع برنولی):  $p_{ij} \equiv p(x_j = 1 | C_i)$

- در صورتی که متغیرها مستقل در نظر گرفته شوند:

$$p(\mathbf{x} | C_i) = \prod_{j=1}^d p_{ij}^{x_j} (1 - p_{ij})^{(1-x_j)}$$

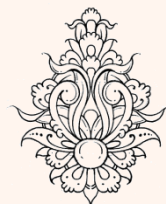
**Naive Bayes'**

$$g_i(\mathbf{x}) = \log p(\mathbf{x} | C_i) + \log P(C_i)$$

$$= \sum_j [x_j \log p_{ij} + (1 - x_j) \log (1 - p_{ij})] + \log P(C_i)$$

**تفمین**

$$\hat{p}_{ij} = \frac{\sum_t x_j^t r_i^t}{\sum_t r_i^t}$$



# خصیصه‌های گسسته (ادامه...)

• در صورتی خصیصه چندمقداری باشد.

•  $x_j \in \{v_1, v_2, \dots, v_{n_j}\}$

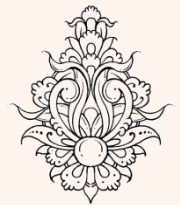
$$p_{ijk} \equiv p(z_{jk} = 1 | C_i) = p(x_j = v_k | C_i)$$

• در صورتی که متغیرها مستقل باشند:

$$p(\mathbf{x} | C_i) = \prod_{j=1}^d \prod_{k=1}^{n_j} p_{ijk}^{z_{jk}}$$

$$g_i(\mathbf{x}) = \sum_j \sum_k z_{jk} \log p_{ijk} + \log P(C_i)$$

$$\hat{p}_{ijk} = \frac{\sum_t z_{jk}^t r_i^t}{\sum_t r_i^t}$$



# رگرسیون خطی

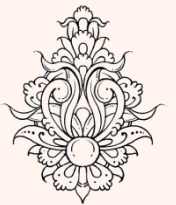
$$g(x^t | w_1, w_0) = w_1 x^t + w_0$$

$$\sum_t r^t = N w_0 + w_1 \sum_t x^t$$

$$\sum_t r^t x^t = w_0 \sum_t x^t + w_1 \sum_t (x^t)^2$$

$$\mathbf{A} = \begin{bmatrix} N & \sum_t x^t \\ \sum_t x^t & \sum_t (x^t)^2 \end{bmatrix} \quad \mathbf{w} = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} \sum_t r^t \\ \sum_t r^t x^t \end{bmatrix}$$

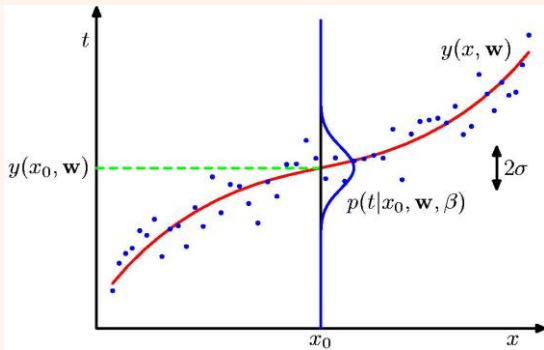
$$\mathbf{w} = \mathbf{A}^{-1} \mathbf{y}$$



# رگرسیون چندجمله‌ای

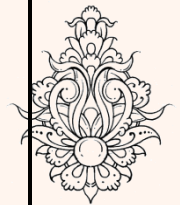
$$g(x^t | w_k, \dots, w_2, w_1, w_0) = w_k (x^t)^k + \dots + w_2 (x^t)^2 + w_1 x^t + w_0$$

$$\mathbf{A} = \begin{bmatrix} N & \sum_t x^t & \sum_t (x^t)^2 & \dots & \sum_t (x^t)^k \\ \sum_t x^t & \dots & \dots & \dots & \sum_t (x^t)^{k+1} \\ \vdots & \dots & \dots & \dots & \vdots \\ \sum_t (x^t)^k & \sum_t (x^t)^{k+1} & \sum_t (x^t)^{k+2} & \dots & \sum_t (x^t)^{2k} \end{bmatrix}$$



$$\mathbf{w} = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_k \end{bmatrix}$$

$$\mathbf{y} = \begin{bmatrix} \sum_t r^t \\ \sum_t r^t x^t \\ \sum_t r^t (x^t)^2 \\ \vdots \\ \sum_t r^t (x^t)^k \end{bmatrix}$$



# رگرسیون چندجمله‌ای

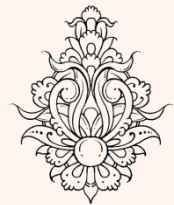
$$\mathbf{A} = \begin{bmatrix} N & \sum_t x^t & \sum_t (x^t)^2 & \dots & \sum_t (x^t)^k \\ \sum_t x^t & & & & \sum_t (x^t)^{k+1} \\ \vdots & & & & \vdots \\ \sum_t (x^t)^k & \sum_t (x^t)^{k+1} & \sum_t (x^t)^{k+2} & \dots & \sum_t (x^t)^{2k} \end{bmatrix}$$

$$\mathbf{y} = \begin{bmatrix} \sum_t r^t \\ \sum_t r^t x^t \\ \sum_t r^t (x^t)^2 \\ \vdots \\ \sum_t r^t (x^t)^k \end{bmatrix}$$

$$\mathbf{A} = (\mathbf{D}^T \mathbf{D}) \quad \mathbf{y} = \mathbf{D}^T \mathbf{r}$$

$$\mathbf{D} = \begin{bmatrix} 1 & x^1 & (x^1)^2 & \dots & (x^1)^k \\ 1 & x^2 & (x^2)^2 & \dots & (x^2)^k \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x^N & (x^N)^2 & \dots & (x^N)^k \end{bmatrix} \quad \mathbf{r} = \begin{bmatrix} r^1 \\ r^2 \\ \vdots \\ r^N \end{bmatrix}$$

$$\mathbf{w} = (\mathbf{D}^T \mathbf{D})^{-1} \mathbf{D}^T \mathbf{r}$$



# رگرسیون چندمتغیره‌ی خطی

*Multivariate linear Regression*

*Multiple Regression*

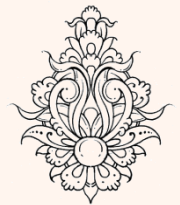
$$r^t = g(x^t | w_0, w_1, \dots, w_d) + \varepsilon$$

$$= w_0 + w_1 x_1^t + w_2 x_2^t + \dots + w_d x_d^t$$

- تابع خطا به صورت زیر به دست می‌آید:

$$E(w_0, w_1, \dots, w_d | \mathcal{X}) = \frac{1}{2} \sum_t [r^t - w_0 - w_1 x_1^t - \dots - w_d x_d^t]^2$$

- مانند آن چه در پیش داشتیم، با مشتق گرفتن، می‌توان ضرایب را به صورت تحلیلی به دست آورد.





# رگرسیون چندمتغیره ی خطی

$$\mathbf{X} = \begin{bmatrix} 1 & x_1^1 & x_2^1 & \dots & x_d^1 \\ 1 & x_1^2 & x_2^2 & \dots & x_d^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_1^N & x_2^N & \dots & x_d^N \end{bmatrix}, \mathbf{w} = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_d \end{bmatrix}, \mathbf{r} = \begin{bmatrix} r^1 \\ r^2 \\ \vdots \\ r^N \end{bmatrix}$$

$$\mathbf{A} = (\mathbf{D}^T \mathbf{D}) \quad \mathbf{y} = \mathbf{D}^T \mathbf{r}$$

$$\mathbf{w} = (\mathbf{D}^T \mathbf{D})^{-1} \mathbf{D}^T \mathbf{r}$$

$$\mathbf{X}^T \mathbf{X} \mathbf{W} = \mathbf{X}^T \mathbf{r} \quad \longrightarrow \quad \mathbf{W} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{r}$$

این مدل شبیه به مدلی است که برای رگرسیون چند جمله‌ای تک متغیره داشتیم.

$$x_1 = x, x_2 = x^2, x_3 = x^3, \dots, x_k = x^k$$

برای رگرسیون چندجمله‌ای و چندمتغیره نیز می‌توانیم به صورت مشابه عمل کنیم:

$$\mathbf{z}_1 = x_1, \mathbf{z}_2 = x_2, \mathbf{z}_3 = x_1^2, \mathbf{z}_4 = x_2^2, \mathbf{z}_5 = x_1 x_2$$

