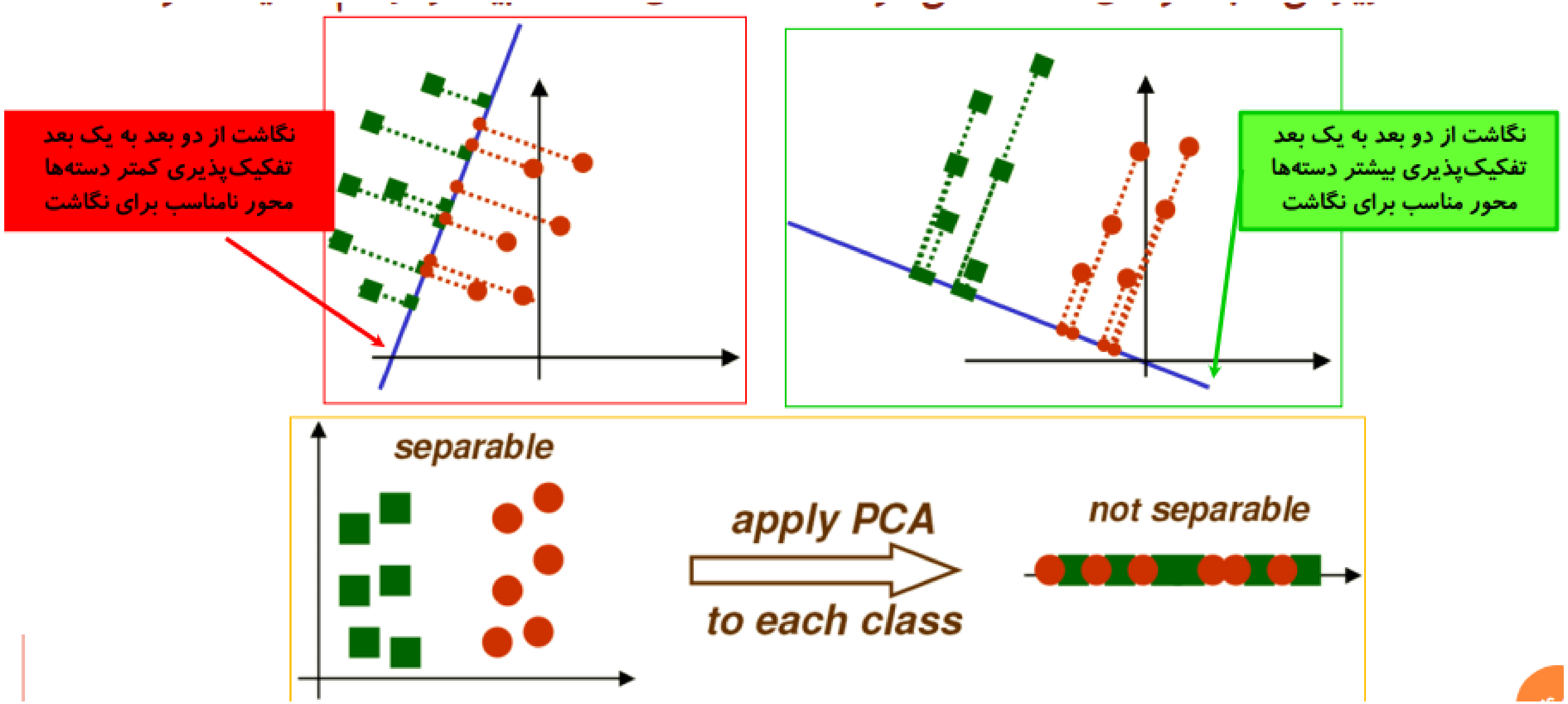
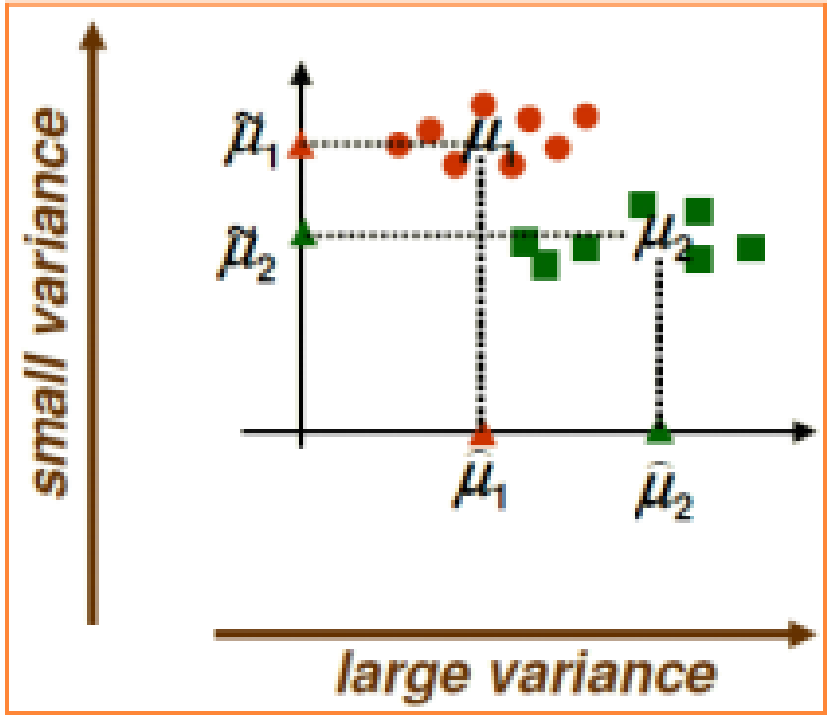
درستمایی بیشینه

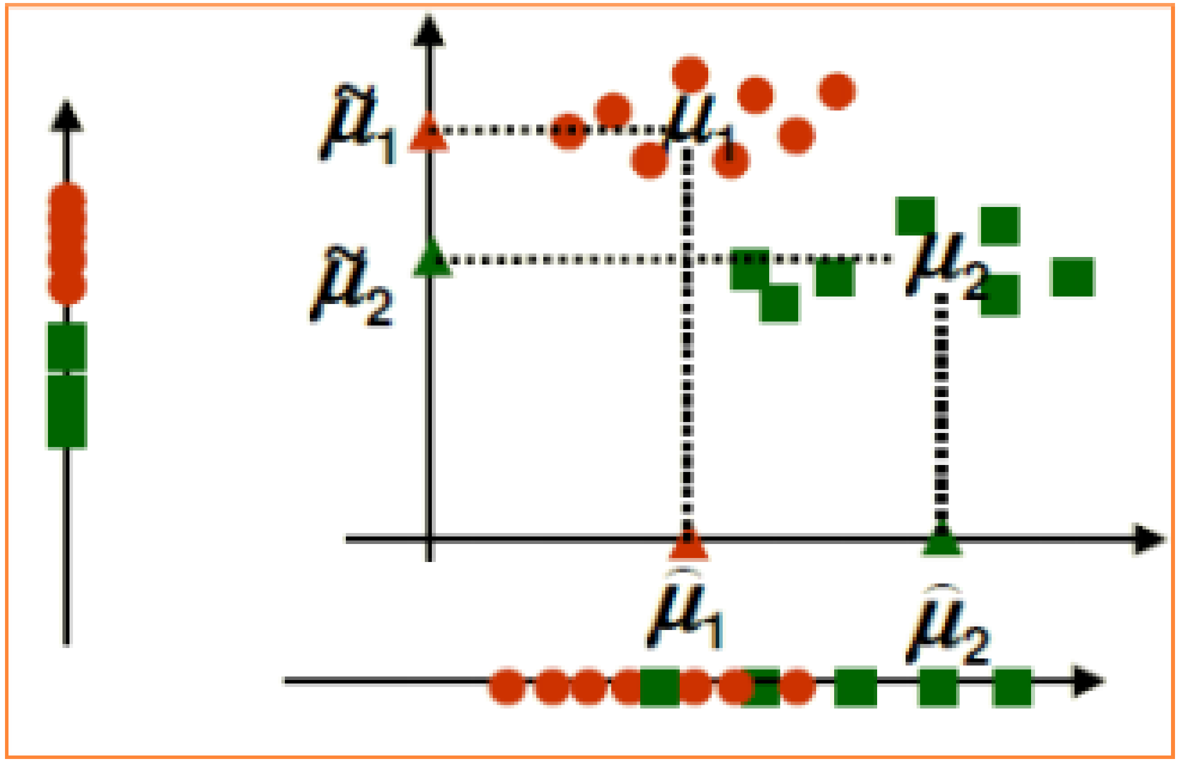
جداساز خطی LDA



هدف

- ویژگی هارو جوری به بعد کمتر نگاشت کنم که با بالاتری میزان تفکیک، کلاس هارو جدا کنیم.





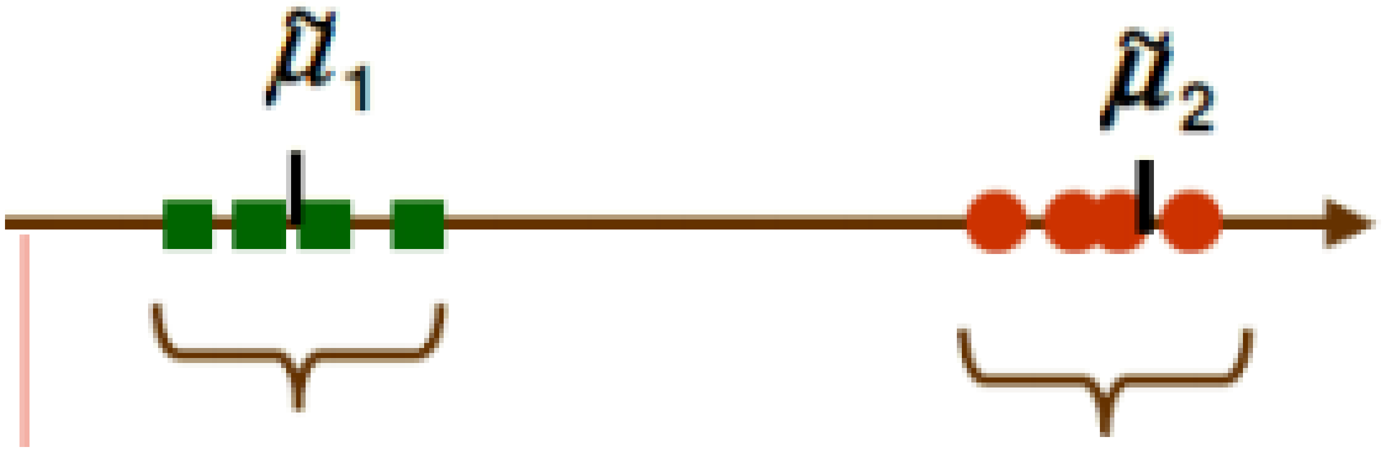
معیار اسکتر یا پراکندگی

- پراکندگی اطراف میانگین(واریانس ضرب در تعداد نمونه ها)

$$S_1^{\,2} = \sum_{x_i \in C_1} (x_i - \mu_1)(x_i - \mu_1)^T \qquad S_2^{\,2} = \sum_{x_i \in C_2} (x_i - \mu_2)(x_i - \mu_2)^T$$

$$S_1'^{\,2} = \sum_{x_i \in C_1} (v^t x_i - \mu_1')(v^t x_i - \mu_1')^T \qquad S_2'^{\,2} = \sum_{x_i \in C_2} (v^t x_i - \mu_2')(v^t x_i - \mu_2')^T$$

هدف: ماکسیمم J



$$\boxed{J(v) = \frac{\left(\mu_1' - \mu_2'\right)^2}{S_1'^2 + S_2'^2}}$$

کمی محاسبات

- J رو تابعی از vرو بنویس

$$S_1'^2 + S_2'^2 = v^t S_1^2 v + v^t S_2^2 v = v^t S_W v$$

$$(\mu_1' - \mu_2')^2 = \left(v^t \mu_1 - v^t \mu_2\right)^2 = v^t (\mu_1 - \mu_2)(\mu_1 - \mu_2)^t v = v^t S_B v$$

$$J(v) = \frac{\left(\mu_1' - \mu_2'\right)^2}{S_1'^2 + S_2'^2} = \frac{v^t S_B v}{v^t S_w v}$$

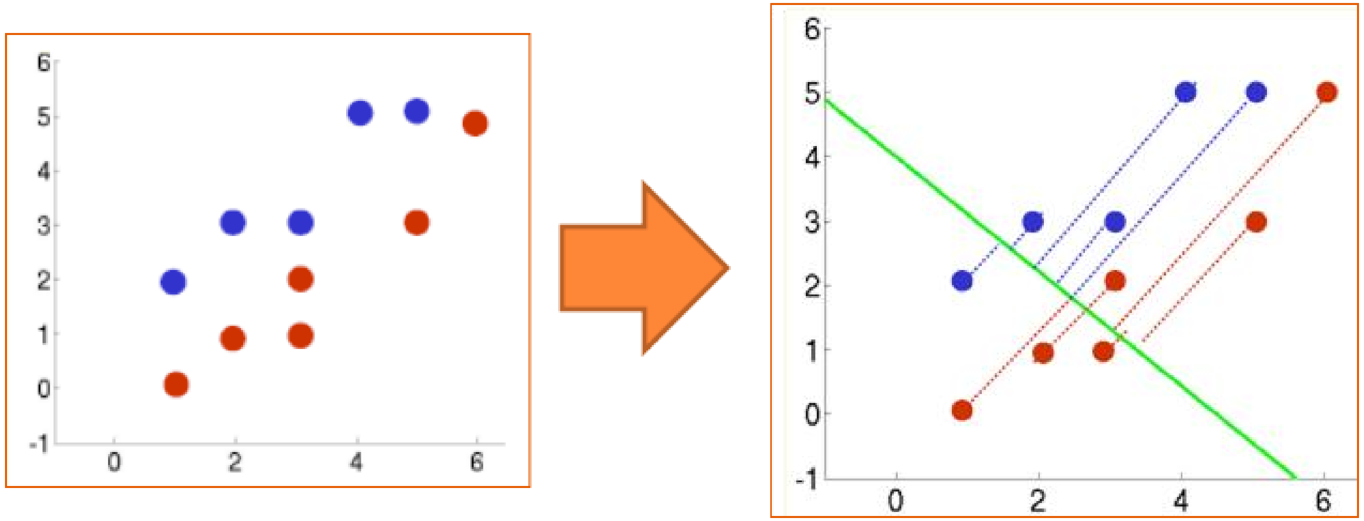
- مشتق بگیر مساوی صفر بزار
- vپیدا میشه.

$$\frac{\partial}{\partial v}J(v)=\frac{\left(\frac{\partial}{\partial v}v^tS_Bv\right)v^tS_Wv-\left(\frac{\partial}{\partial v}v^tS_Wv\right)v^tS_Bv}{\left(v^tS_wv\right)^2}=\frac{(2S_Bv)v^tS_Wv-(2S_Wv)v^tS_Bv}{\left(v^tS_wv\right)^2}$$

- اندازه v مهم نیست، پس

$$v=S_W^{-1}\left(\mu_1-\mu_2\right)$$

- نکته: محل دقیق خط نگاشت مهم نیست، جهت آن مهم است



١

QDAیا MDA

- کلاسبندی nنمونه در C کلاس
- گام اول: محاسبه میانگین هر دسته و کل

$$\mu_i = \frac{1}{n_i} \sum_{x_i \in C_i} x_i$$

$$\mu = \frac{1}{n} \sum_{All\ x_i} x_i = \frac{1}{n} \sum_{i=1}^{C} n_i \mu_i$$

- گام دوم: محاسبه ماتریس پراکندگی درون دسته ای و بین دسته ای:

$$S_B = \sum_{i=1}^{C} n_i (\mu_i - \mu)(\mu_i - \mu)^t \qquad S_W = \sum_{i=1}^{C} S_i = \sum_{i=1}^{C} \sum_{x_k \in C_i} (x_k - \mu_i)(x_k - \mu_i)^t$$

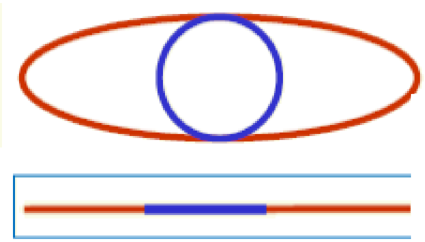
تابع هدف(محاسبه بردار نگاشت v)

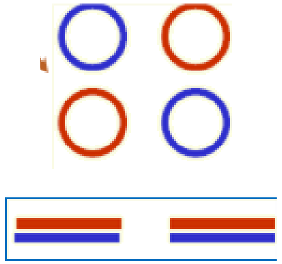
$$J(v) = \frac{\det\left(v^t S_B v\right)}{\det\left(v^t S_W v\right)}$$

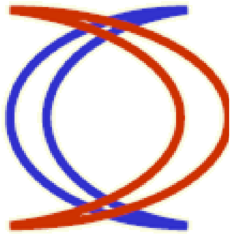
$$S_B v = \alpha S_W v$$

عیوب

- کاهش بعد حداکثر تا یدونه کمتر تعداد کلاسها
- اگر میانگین ها برابر باشند و تفکیک فقط به واریانسها وابسته باشه گاها مشکل میخوریم
- اگر تابع هدفمون خیلی زیاد بشه یعنی اختلاف میانگین ها بالا باشه، داده ها بعد نگاشت ، اورلپ زیادی دارند.





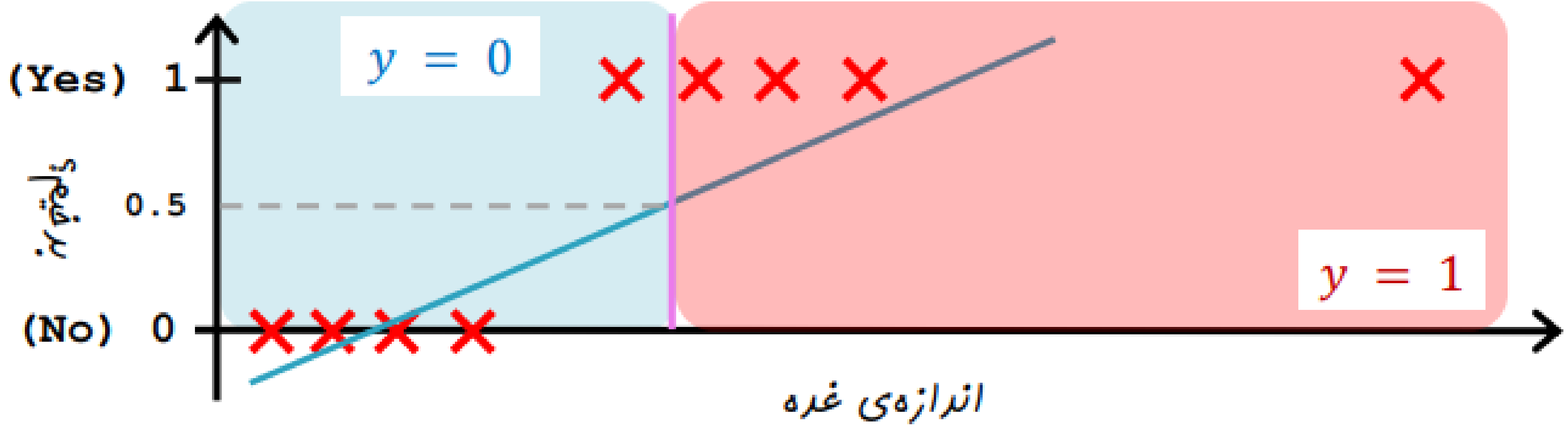


درستنمایی بیشینه

- به زبان ساده:
  - پارامترهای مساله رو جوری یاد بگیرم که بتونم با بالا ترین احتمال ممکن، نمونه هامو به درستی کلاس بندی کنم.
  - پارامترهام چیا هستن؟
    - بسته به مساله متفاوته ولی مثلا میتونه میانگین و واریانس باشه یا مثلا اگر همون ضرایبی که توی مساله چکرز داشتیم.

- اگه از دید یادگیری ماشین بخوام بیشتر بگم:
  - فرضیه تابع هدفمو ازش یه تفسیر احتمالاتی کنم.
  - یعنی مثلا تو مساله چکرز، بگم چقدر با این شرایط فعلی برد، احتمال داره برم به برد مورد نظر بعدی. همونجا هم ما از امتیاز حالات استفاده میکردیم. اینجا همونو اسمشو بزار احتمال و باهاش با رویکرد احتمالاتی برخورد کن!

دسته بندی



- هدف.

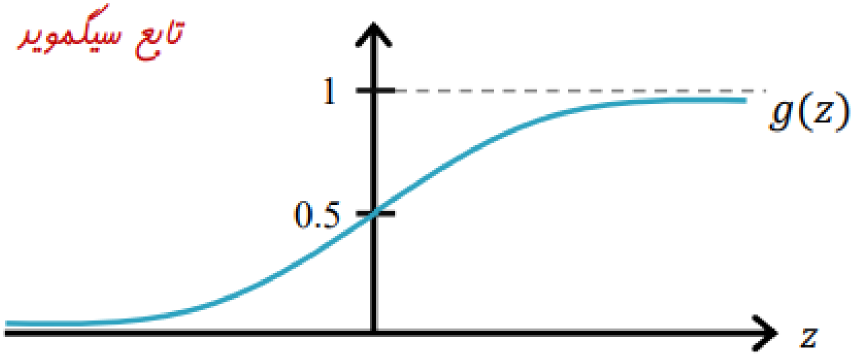
$$0 \leq h_\theta(x) \leq 1$$

- فرضیه.

$$h_\theta(x) = g(\theta^T x)$$

$$g(z) = \frac{1}{1 + e^{-z}}$$

$$g(\theta^T x) = \frac{1}{1 + e^{-\theta^T x}}$$



تفسیر احتمالاتی

برنولی

- تفسیر احتمالاتی فرضیه.

$$p(y=1|x;\theta) = h_\theta(x)$$

$$p(y=0|x;\theta) = 1 - h_\theta(x)$$

$$p(y|x;\theta) = h_\theta(x)^y \left(1 - h_\theta(x)\right)^{1-y}$$

تخمین بیشترین درست‌نمایی

- تابع درست‌نمایی.

$$L(\theta) = p(Y|X;\theta) = \prod_{i=1}^{m} p\left(y^{(i)}|x^{(i)};\theta\right)$$

$$= \prod_{i=1}^{m} h_\theta\left(x^{(i)}\right)^{y^{(i)}} \left(1 - h_\theta\left(x^{(i)}\right)\right)^{1-y^{(i)}}$$

حاصلضرب تک تک احتمالات -------> خیلییی کوچیک میشه پسسسس: ماکزیمم لگاریتمو بگیر

میخوام احتمال دسته بندی درست تک تک داده هارو بیشینه کنم

- لگاریتم تابع درست‌نمایی.

$$
\begin{aligned}
l(\theta) &= \log L(\theta) = \log \prod_{i=1}^{m} h_\theta(x^{(i)})^{y^{(i)}} \left(1 - h_\theta(x^{(i)})\right)^{1-y^{(i)}} \\
&= \sum_{i=1}^{m} \log\left(h_\theta(x^{(i)})^{y^{(i)}} \left(1 - h_\theta(x^{(i)})\right)^{1-y^{(i)}}\right) \\
&= \sum_{i=1}^{m} y^{(i)} \log h_\theta(x^{(i)}) + (1 - y^{(i)}) \log\left(1 - h_\theta(x^{(i)})\right)
\end{aligned}
$$

□ تابع هزینه.

$$J(\theta) = -l(\theta) = \sum_{i=1}^{m} -y^{(i)} \log h_\theta\left(x^{(i)}\right) - \left(1 - y^{(i)}\right) \log\left(1 - h_\theta\left(x^{(i)}\right)\right)$$

- خب! حالا من با مشتق گرفتن و صفر قرار دادن تابع درستنمایی بیشینه، پارامترهامو پیدا میکنم. این پارامترها، هزینه یادگیری منو، خطای تصمیم گیری منو، کمینه میکنه. مثلاا، توی چکرز، اگر پارامترهای مثلا یه مقدار خاص باشن بهترین فرضیه ممکنو دارم و برد احتمالیش بالا میره. یا مثلا توی مسایل احتمالاتی و توزیع های نرمال، میتونم اگه پارامترهایی مثل واراینسو اگه مد نظرمه به بهترین شکل ممکن در جهت ارضای هدف مساله (مرز تصمیم) پیدا کنم.

- در واقع من از روی نمونه کوچکی داده، تابع رفتار کلی جامعه رو حدس زدم

تست کنکور96

- دو سکه داریم که سکه یک با احتمال θ و سکه دوم با احتمال 2θ شیر می آید. این دو سکه را چندین بار پرتاب میکنیم و نتیجه به صورت زیر است. درست نمایی بیشینه برای پارامتر θ کدام است؟

سکه	1	2	2	2	2
نتیجه	شیر	خط	خط	خط	شیر

- $T = \theta(1-2\theta)(1-2\theta)(1-2\theta)2\theta = 2\theta^2(1-2\theta)^3$
- $\log T = \log(2\theta^2(1-2\theta)^3) = 2\log 2\theta + 3\log(1-2\theta)$
- $\partial T/\partial\theta = 2/2\theta + 3/1-2\theta = 0$
- $\theta = 1/5$