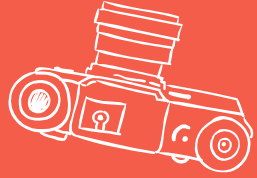
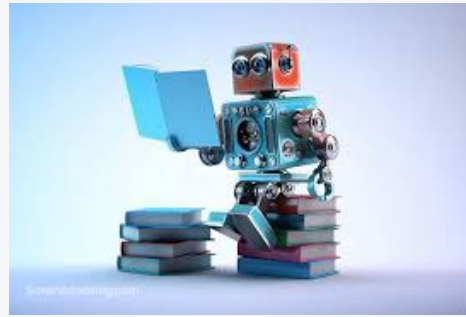


به نام خدا



یادگیری ماشین





یادگیری ماشین

آرش عبدی هجراندوست

arash.abdi.hejrandoost@gmail.com

دانشگاه علم و صنعت

دانشکده مهندسی کامپیوتر

نیم سال اول ۱۴۰۱-۱۴۰۲

TA

✗ تدریس‌یار: آقای آرمین توکلی
seniort 1378@gmail. com ○

آزمون اطمینان Significant Test

× فرضیه صفر H_0 Null Hypothesis

× رد یا قبول با درجه اطمینان دلخواه

× مثلاً در هرس گره های درخت تصمیم

○ معیار کوچک بودن information gain

○ تا چه مدی کوچک؟

○ فرض: ویژگی انتخاب شده هیچ تمایزی ایجاد نمی‌کند/ هیچ الگویی ندارد

■ فرضیه صفر

■ بررسی میزان درست/غلط بودن این فرضیه

هرس کای ۲ - X^2

❌ فرضیه صفر:

- ویژگی انتخاب شده هیچ تمایزی ایجاد نمی‌کند.
- محاسبه میزان احتمال بر اساس توزیع تصادفی
- اندازه نمونه ها: $v = n + p$

$$\hat{p}_k = p \times \frac{p_k + n_k}{p + n}, \quad \hat{n}_k = n \times \frac{p_k + n_k}{p + n}$$

○ میزان انحراف از فرضیه:

$$\Delta = \sum_{k=1}^d \frac{(p_k - \hat{p}_k)^2}{\hat{p}_k} + \frac{(n_k - \hat{n}_k)^2}{\hat{n}_k}$$

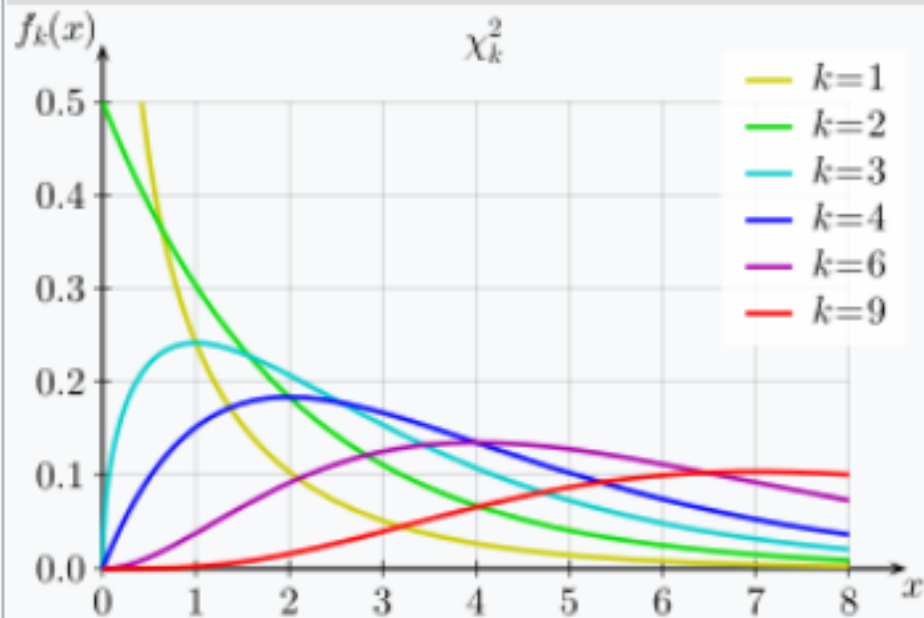
○ توزیع فوق از توزیع X^2 با درجه $d - 1$ است و میتوان بر اساس نمونه موجود میزان اطمینان را محاسبه کرد.

هرس کای ۲ - χ^2

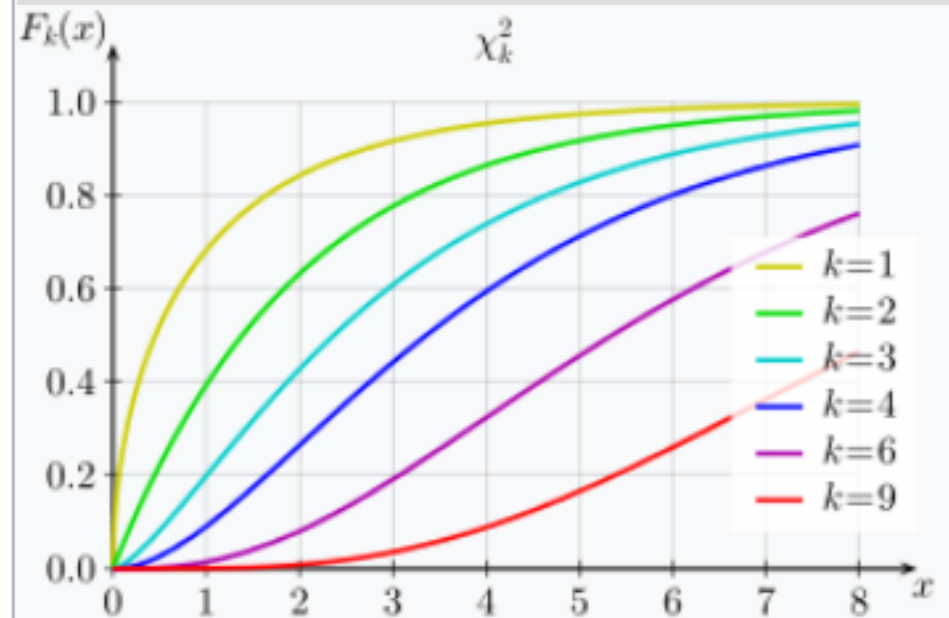
✗ In probability theory and statistics, the **chi-squared distribution** (also **chi-square** or χ^2 -distribution) with k degrees of freedom is the distribution of a sum of the squares of k independent standard normal random variables.

chi-squared

Probability density function



Cumulative distribution function



Degrees of freedom (df)	χ^2 value ^[20]										
1	0.004	0.02	0.06	0.15	0.46	1.07	1.64	2.71	3.84	6.63	10.83
2	0.10	0.21	0.45	0.71	1.39	2.41	3.22	4.61	5.99	9.21	13.82
3	0.35	0.58	1.01	1.42	2.37	3.66	4.64	6.25	7.81	11.34	16.27
4	0.71	1.06	1.65	2.20	3.36	4.88	5.99	7.78	9.49	13.28	18.47
5	1.14	1.61	2.34	3.00	4.35	6.06	7.29	9.24	11.07	15.09	20.52
6	1.63	2.20	3.07	3.83	5.35	7.23	8.56	10.64	12.59	16.81	22.46
7	2.17	2.83	3.82	4.67	6.35	8.38	9.80	12.02	14.07	18.48	24.32
8	2.73	3.49	4.59	5.53	7.34	9.52	11.03	13.36	15.51	20.09	26.12
9	3.32	4.17	5.38	6.39	8.34	10.66	12.24	14.68	16.92	21.67	27.88
10	3.94	4.87	6.18	7.27	9.34	11.78	13.44	15.99	18.31	23.21	29.59
<i>p</i> -value (probability)	0.95	0.90	0.80	0.70	0.50	0.30	0.20	0.10	0.05	0.01	0.001

در مثال رستوران: ○

ویژگی type ■

۴ مقدار ← درجه ۳ ■

● در واقع ۳ متغیر تصادفی وجود دارد (چهارمی؟)

■ $\Delta = 7.82$ یا بیشتر ← (د فرضیه صفر با درجه اطمینان ۹۵٪) (۵٪)

■ $\Delta = 11.35$ یا بیشتر ← (د فرضیه صفر با درجه اطمینان ۹۹٪) (۱٪)

■ مقادیر کمتر ← پذیرش فرضیه بی ربط بودن ویژگی

چند چالش در درخت تصمیم

Missing Data ✗

○ با وجود درخت، داده تست بدون یک ویژگی مشخص را چگونه دسته بندی کنیم؟

○ چگونه آنتروپی و Gain محاسبه کنیم؟

اگر ویژگی دارای مقادیر گسسته بسیار زیاد بود؟ ✗

○ کد شهر

ناپایداری درخت ✗

○ دیدن یک نمونه جدید = احتمال تغییر ریشه درخت و ...

انتخاب مدل و بهینه سازی

× هدف کلان یادگیری

× فرض iid برای نمونه ها

○ independent and identically distributed

× Error rate

× Hyperparameters

× تقسیم داده ها:

○ Training set

○ Validation set

○ Test set

× K- fold cross- validation

× leave- one- out cross- validation - LOOCV ← $K=n$

انتخاب مدل و بهینه سازی

✗ انتخاب مدل

✗ انتخاب فضای مدل

✗ ...

○ چند جمله ای

○ درجه ۲

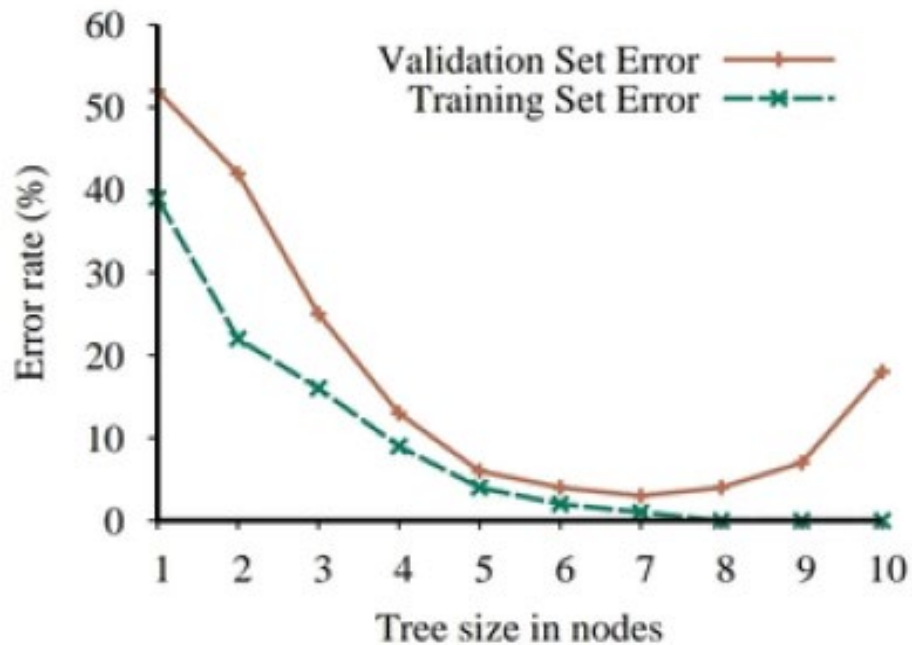
○ با ضرایب مشخص

✗ انتخاب فضای مدل:

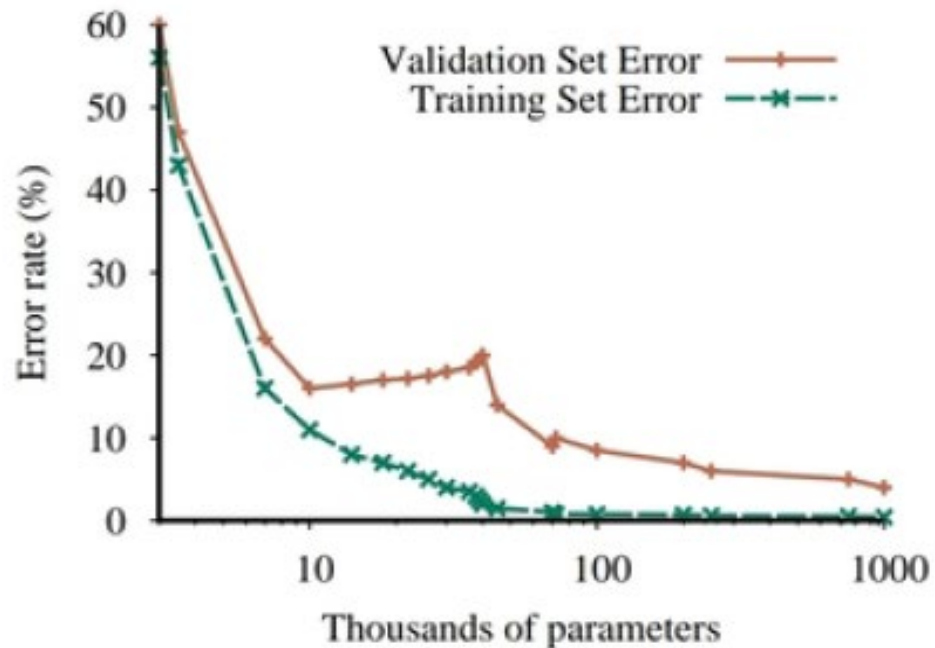
○ Subjective

○ objective

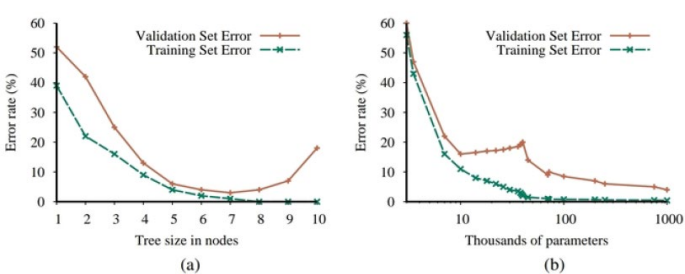
انتخاب مدل validation



(a)



(b)



ظرفیت مدل

✗ ظرفیت درخت تصمیم

○ برگ به تعداد نمونه ها

✗ تطابق کامل با داده های آموزشی:

○ Interpolation

○ Memorization

✗ افزودن ظرفیت به مدل (مثلا شبکه عصبی)

○ توابع کانید بیشتر

○ احتمال یافتن تابع مناسب بیشتر

Error Rates

✗ آیا درصد خطا برای ارزیابی کارایی کافی است؟

○ لابد نه؟

✗ دسته بندی Spam

○ فضای ۱٪ خوب است؟

✗ Loss- خسارت θ برعکس utility

$$L(x, y, \hat{y}) = \text{Utility}(\text{result of using } y \text{ given input } x) - \text{Utility}(\text{result of using } \hat{y} \text{ given input } x)$$

✗ وزن دادن به نوع خطا

○ بیشتر از دو کلاس

✗ خروجی پیوسته:

تعریف L بر اساس نرم L_p ○

L_2 ○

L_1 ○

L_0 ○

Generalization Loss

$$GenLoss_L(h) = \sum_{(x,y) \in \mathcal{E}} L(y, h(x)) P(x, y)$$

$$h^* = \operatorname{argmin}_{h \in \mathcal{H}} GenLoss_L(h)$$

$$EmpLoss_{L,E}(h) = \sum_{(x,y) \in E} L(y, h(x)) \frac{1}{N}$$

$$\hat{h}^* = \operatorname{argmin}_{h \in \mathcal{H}} EmpLoss_{L,E}(h)$$

علل وجود خطا: ❌

○ فضای محدود (فرضیه)

○ واریانس

■ تأخیر نسبت به دیتاست

○ Noise

○ بار محاسباتی

■ جستجوی همه فرضیه ها ممکن نیست

Regularization

هدف از Cross - Validation ❌

○ جلوگیری از بیش برآزش

❌ راه جایگزین- جریمه کردن پیچیدگی:

$$Cost(h) = EmpLoss(h) + \lambda Complexity(h)$$

$$h^* = argmin Cost(h)$$

❌ معیاری پیچیدگی؟

❌ L_2 برای فضای چندجمله ای: مثلاً

○ جلوگیری از خروجی های شدیداً پرت!

Feature Selection

× راه دیگری برای تامین سادگی در مدل

× هرس X^2

× روشهای عمومی انتخاب ویژگی

تنظیم پارامترهای آزاد

✗ یافتن مقدار صحیح پارامترها با Cross - Validation

- بیش از یک پارامتر
- مقادیر زیاد یک پارامتر (پیوسته)

✗ Hand - tuning

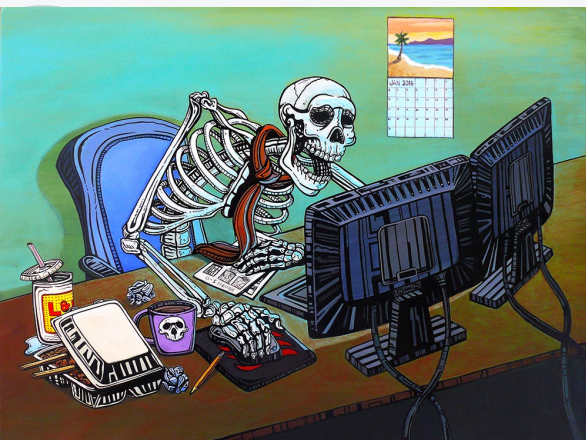
- تنظیم دستی
- کیلویی!!
- ایراد:

■ باید بالا سر سیستم بود ...

● و ماند!

○ و مُرد!!

■ لزوماً به بهترین نمی‌رسیم! (به اندازه کافی خوب)



تنظیم پارامترهای آزاد

Grid search ✗

○ امکان موازی سازی اجراها

Random Sampling ✗

○ چشم بسته تا وقتی وقت هست

بهینه سازی تابع یادگیری ✗

○ یافتن تابعی با ورودی پارامترها و خروجی دقت (loss)

○ بعد از هر بار محاسبه برای یک سری پارامتر، فرضیات راجع به تابع هدف به روز شود

آموزش مبتنی بر جمعیت - population-based training ✗

○ الگوریتم های ژنتیک - تکاملی



با تشکر

