Métricas de qualidade de ajuste

No mundo dos dados, nunca devemos ter um modelo preferido. O ideal é sempre testar tipos de algorítimos diferentes e escolher o melhor. Este e-book irá gerar *insights* e orientá-lo na escolha da melhor métrica para definir seu modelo campeão.

Data: fevereiro/2019

Divirta-se aprendendo de verdade!

www.asn.rocks

Divirta-se aprendendo de verdade!

Adriana Silva

Sou uma eterna apaixonada por ensinar e compartilhar o pouco que sei. Amo estatística e tenho como objetivo de vida mostrar para o mundo o quanto isso é sexy, incrível e esclarecedor. Tenho certeza que a ciência irá mudar nossas vidas! Quer me ajudar nessa missão?

/in/adrianamms dri@asn.rocks www.asn.rocks

SUMÁRIO

01 INTRODUÇÃO

02 AS 3 UTILIDADES DE UM MODELO

03 MÉTRICAS MAIS CONHECIDAS

04 APLICAÇÃO DE NEGÓCIO

05 CONCLUSÃO

01.

INTRODUÇÃO

INTRODUÇÃO

Todo cientista de dados tem inúmeros desafios durante suas atividades diárias. E uma das atividades que é mais desafiadora é “bater o martelo” na escolha do melhor modelo.

No momento em que vários modelos foram feitos e estão todos corretos, temos que escolher um e, aí vem a pergunta: “Como escolher o melhor modelo?”.

Este e-book tem o objetivo de discutir as métricas e explicar como elas funcionam. Apresentarei 3 métricas mais famosas, mas lembre-se que existem várias outras.

No entanto, o tema desse e-book também se relaciona a alguns outros temas que já discuti em outro e-book e vídeo. Não deixe de acessar meu site e procurar pelo e-book “Como utilizar um modelo” e também não perca a oportunidade de assistir a vídeo aula sobre “Overfitting”. Eles irão ajudar caso alguns conceitos discutidos aqui soem novos ou não fiquem claros.

Para melhor absorção desse material, presumo que você já saiba para que serve a partição de dados: treinamento, validação e teste e que entende muito bem a essência disso no contexto de modelagem, ok?

Vamos nessa....

02.

AS 3 UTILIDADES DE UM MODELO

AS 3 UTILIDADES DE UM MODELO

Como já mencionei no outro e-book "Como utilizar um modelo", vou apenas relembrar os propósitos que nos levam a modelar:

1) DECISÃO

Nete ponto, temos modelos que são capazes de predizer um fato com precisão, ou seja, que tem uma taxa de resposta alta. O que isso quer dizer? Quer dizer que o modelo criado será utilizado, sem restrições, e alguma ação será tomada em todos os indivíduos que forem classificados como o evento em questão.

2) ORDENAÇÃO

Já neste caso, temos modelos que são capazes de ordenar a probabilidade de um evento acontecer. O que isso quer dizer? Quer dizer que o modelo criado será utilizado, com algumas restrições de verba por exemplo, e alguma ação será tomada apenas nos indivíduos mais propensos a cometerem o evento em questão. Repare que neste caso a ordem importa, diferente do anterior que se importava com a decisão tomada.

3) ESTIMAÇÃO

No terceiro e último caso, temos modelos que são precisos suficientes para estimar a probabilidade de algo acontecer sem muito erro, ou seja, a estimativa deve estar bem próxima da realidade. O que isso quer dizer? Quer dizer que o modelo criado será utilizado com o objetivo de multiplicar essa probabilidade por algum outro valor, por isso a tamanha importância do menor erro possível.

E isso é extremamente importante estar claro, pois é a partir disso que vamos escolher a métrica que faz mais sentido para o meu problema de negócio. Se isso pareceu estranho, não deixe de ler o e-book mencionado. Nele existem exemplos de aplicação e propósito para cada um dos 3 motivos.

03.

MÉTRICAS MAIS CONHECIDAS

MÉTRICAS MAIS CONHECIDAS

Existem diversas métricas e todas elas tem algum propósito (ninguém inventa algo sem ter alguma razão, certo?).

O importante do nosso papel é descobrir essa razão e tirar o melhor proveito possível desse conteúdo para você poder escolher o que se encaixa melhor no seu problema.

É bizarro o tanto de coisa que temos que saber para ter um bom resultado, né? Vou tentar colaborar um pouco com isso...

Vou descrever, da melhor forma que eu conseguir, uma série de métricas conforme as 3 utilidades de um modelo.

Como assim?

Para cada propósito de modelo temos métricas que podem nos contribuir. Explicarei essas:

1) DECISÃO
   - Misclassification
2) ORDENAÇÃO
   - Lift
3) ESTIMAÇÃO
   - ASE

Vou dar a explicação de apenas 3 métricas, mas tenha ciência de que existem diversas outras, que vou citar nomes aqui para que você possa pesquisar, ok?

Vamos começar pela mais simples, o *misclassification*. Esta métrica é, sem dúvida, a mais tranquila de ser calculada.

O foco desta métrica é calcular o percentual de erro de previsão... Com objetivo de decisão.

Deixe me detalhar um pouco mais utilizando um exemplo.

Vamos imaginar que estamos trabalhando em um modelo, onde nosso foco é saber se a pessoa vai engordar ou não vai engordar baseado em seus hábitos (eu coloquei a fita métrica na página anterior e me lembrei que estou engordando... Huahuahua).

Bom, então dado os indivíduos que pegamos para estudar, sabemos os que engordaram e os que não engordaram. Baseado nas variáveis independentes criamos um modelo que estima a probabilidade dele engordar (evento de interesse). Baseado nessa probabilidade, definimos um ponto de corte (chamado de cutoff... E isso pode ser um novo e-book... vou pensar sobre isso...) de 0,5. Então, todos que estiverem com a probabilidade maior ou igual a 0,5 chamaremos de "engorda" e os ques estiverem com uma probabilidade menor que isso, chamaremos de "não engorda".

Sendo assim, temos a previsão do modelo e o fato real que aconteceu nos nossos dados de treinamento e validação.

Então poderíamos criar a famosa e tão incrível matriz de confusão:

		Decisão do Modelo	
		Engorda	Não Engorda
Vida real	Engorda	a	b
	Não Engorda	c	d

As letrinhas a, b, c e d representam a quantidade de linhas que:

a) A verdade era "engorda" e meu modelo chamou de "engorda" ✓

b) A verdade era "engorda" e meu modelo chamou de "não engorda" ✗

c) A verdade era "não engorda" e meu modelo chamou de "engorda" ✗

d) A verdade era "não engorda" e meu modelo chamou de "não engorda" ✓

Bom, se quero saber em quantos porcento eu errei, basta somar b e c e dividir pela soma de todos, certo?

Este é o tão famoso *misclassification*. Uma medida bem genérica. Tão genérica que eu, particulamente, não gosto dela. No entanto minha opinião não conta tanto, visto que vejo inúmeras pessoas decidindo apenas com essa informação.

No meu critério, as métricas *precision* e *recall* são bem mais sexy que esta e vem da mesma matriz de confusão.

Uma métrica mega animal (essa eu gosto de verdade, amor incondicional porque a acho genial) é o Lift (alguns chamam de Index também).

Estou ficando emocionada... essa é demais, deixe-me lhe contar.

Se seu propósito é ordenação, isso quer dizer que você irá priorizar alguém (ou algo) que receber a probabilidade mais alta no seu modelo. Vamos imaginar que preciso priorizar uma lista de investimentos pela probabilidade dele dar alta (uma vez que meu dinheiro é limitado e não posso colocar em todos os fundos).

Então, para cada fundo teremos a probabilidade estimada, certo?

Agora, por favor, comece a imaginar aí no seu cérebro o que vou descrever aqui. Imagine uma tabela, onde temos o Y (alta=1 ou baixa=0 do fundo em questão), as variáveis X´s e o y estimado (a famosa probabilidade estimada pelo modelo).

Se nosso objetivo é ordenação, então, por favor ordene esta tabela no seu cérebro pela probabilidade estimada, da maior para a menor.

Beleza, guarda essa tabela aí, por favor! Vamos pensar numa coisa. Na tabela existe a variável Y real (alta ou baixa do fundo em questão), certo? Nela conseguimos calcular a taxa de resposta (que nada mais é que o número de eventos de interesse dividido pelo número total de linhas na tabela). Vamos imaginar que essa taxa seja 0,5 na base toda.

Bom, agora vamos voltar para aquela tabela ordenada pela probabilidade estimada, que você guardou aí no seu cérebro. Se liga, se a proposta é ordenar, é porque não é tudo que me interessa e sim um percentual ou volume pequeno dos primeiros com maiores probabilidades, certo?

Então, volta a tabela ordenada aí pro seu campo de visão imaginária. Agora, pegue os primeiros 10% das linhas dessa tabela (que está ordenada pela probabilidade estimada, certo?). O que você espera da taxa de resposta dos primeiros 10% das linhas? Que ela seja maior, menor ou igual a que encontramos na tabela toda?

Pô, se a base está ordenada pela probabilidade do meu modelo πk, então espero que ele tenha colocado os 1´s (uns) reais (lembrando que os 1´s aqui neste caso são os investimentos em alta) lá pra cima, certo? Logo, espero fortemente que a taxa estimada nos primeiros 10% das linhas seja bem maior que a média total (percentual de 1´s na base toda).

Isso quer dizer que o modelo ordenou bem, fez o papel dele, subiu os 1´s. Então que tal medir essa diferença? Quanto a taxa de resposta nos primeiros 10% é maior que a taxa média?

Jesus Cristo, olha que genial e simples! Suponha que a taxa de resposta nos primeiros 10% seja de 0,9. Então o Lift nesse caso será 0,9/0,5 = 1,8.

Como interpretamos isso? Essa é a parte que eu mais gosto... Esse número quer dizer que o cientista de dados, dono do modelo, obteve uma taxa de resposta 1,8 vezes a taxa obtida em uma seleção aleatória, ou seja, causou um incremento de 0,8 no resultado.

Caraca, falar um negócio desse para meu chefe dá até orgulho, né? Imagina quando esse lift dá algo como 4, então! "Se liga chefão, eu sou 3 vezes melhor que tudo que você fez até hoje (assumindo que ele selecionava indivíduos aleatoriamente), chuuupa!!!" huahuahuahua, claro que nunca falaremos um negócio desses, visto que temos um grande foco em comunicação, mas que dá aquela vontadinha de vez em quando... dá! Huahuahua

Enfim, esta é uma métrica calculada por "pedaços" da base. Eu aqui no meu caso escolhi os percentis (10%). No entanto você pode calcular da forma que fizer mais sentido para sua história, ok?

É uma métrica linda (LOVE pra ela) e super sexy, útil e que te permite falar um "chuuupa" pro chefão! Ahuahuauhahua

14

A última métrica, e não menos importante, que falaremos é o ASE... Cujo nome completo é Average Score Error.

Esta é uma medida que vai mensurar o quanto estamos distante da verdade. Bom, imagine que nosso evento resposta seja 0 ou 1. O que vamos estimar é uma probabilidade desse evento acontecer, então teremos valores ENTRE 0 e 1, certinho?

Bem, pensa aqui comigo... Se a linha dos dados que estou estimando é 0 (não aconteceu) e eu estimo uma probabilidade 0,1... Eu estou bem na fita, certo? (0 – 0,1 = -0,1) Mas e se eu tivesse estimado algo como 0,8... Puts... Errei grande... Tipo 0,8 (= 0 - 0,8).

Sacou a ideia? Se fosse 1 e eu estimasse 0,1... Eu teria errado tanto quanto 0,9 (= 1 - 0,1)... E se tivesse estimado 0,89, eu teria mandado bem (1 - 0,89 = 0,11). Pegou?

Então, o ASE nada mais é do que a média dos erros linha a linha da tabela. Quanto menor esse número, melhor será a qualidade de ajuste do meu modelo. Esta é uma medida muito mais rica para comparação. Interpretá-la para um único modelo não vai levá-lo a lugar nenhum, pois ela não tem uma taxa média de referência, como as outras tem... Então, não tiramos muito valor desse resultado sozinho... Mas quando temos vários, sabemos quem teve melhor desempenho, sacou?

Aqui eu mostrei um exemplo usando-a para uma variável resposta binária, mas, sem dúvida, essa métrica é muito útil para variáveis respostas numéricas também. Inclusive, dentro das 3 métricas que expliquei, seria a única que faz sentido no caso de previsão de número e não de ocorrência de eventos.

04.

APLICAÇÃO DE NEGÓCIO

APLICAÇÃO DE NEGÓCIO

Aplicações e histórias todo mundo gosta, né? Hoje li uma reportagem sobre fraude, então estou com isso na cabeça... Vamos usar isso de exemplo?

Bom, se você respondeu não, perdeu... Eu já estou escrevendo sobre fraude! Vamos imaginar que somos um varejo imenso, que vendemos produtos para gente rica e esnobe, estranha e esquisita.... Sabe? Aquelas que adoram comprar coisas caras...

Como todo e qualquer negócio, estamos expostos à pessoas mal intencionadas e que querem se beneficar de alguma forma. Sendo assim, nossa loja não é diferente das outras e, de vez em quando, temos casos de transações fraudulentas... e ficamos com um preju. =[

Nesse cenário, vamos contratar os "nerds" para construírem modelos para nós, com o objetivo de prever se alguma transação é fraudulenta ou não.

Repare que a variável resposta é a mesma, que as técnicas que o cara vai usar, podem ser as de sempre.... No entanto.. Se esse cara (o "nerd") não for esperto, pode não me oferecer o melhor modelo para cada um dos cenários que vamos criar...

Para ficar mais claro... Vamos continuar com a nossa história do varejo...

Cenário 1

Neste cenário, estamos interessados em saber qual transação é ou não fraude, pois vamos bloquear TODAS essas transações e averiguar a veracidade dos dados inseridos para compra. Estamos malucos para bloquear fraude, então vamos investir muito nisso.

Veja isso aí... Existe alguma restrição? Não... Ainda estão investindo com a sede de acabar com as fraudes... Veja... O mais importante aqui é que meu modelo saiba DECIDIR... Então neste caso, espero fortemente que o "nerd" escolha uma métrica como *misclassification*, *precision*, *recall*, *F1 Score* ou algo do tipo para escolher o melhor modelo.

Cenário 2

Agora, estamos interessados na mesma coisa que é saber quais são as transações fraudulentas, mas só atuaremos nas que nossa mesa (pessoas que irão investigar a possível fraude) conseguir atuar. Normalmente esta capacidade é de 10 pesquisas por dia (nessa minha história, isso não é um limite de mercado).

Olha lá, rapaz...neste caso o objetivo é o mesmo, no entanto, existe uma restrição sobre o alcance de execução em cima do modelo. Com isso, espera-se que a métrica de escolha do modelo seja algo como o *Lift* ou *KS*, por exemplo.

Cenário 3

Continuamos interessados em saber a transação fraudulenta e também continuamos com a restrição da nossa mesa, mas agora queremos focar nas transações que causarão maior preju para nós....

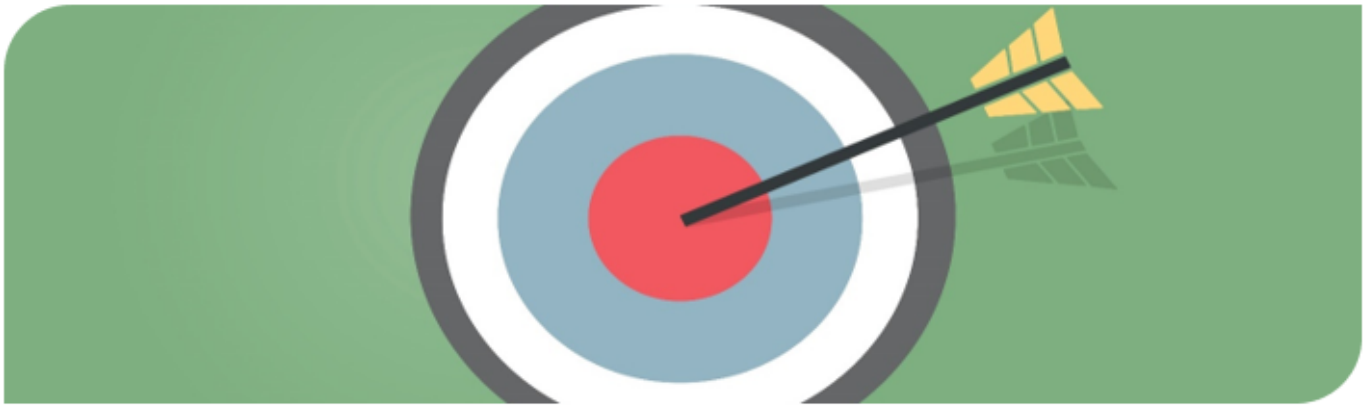
Olha que sensacional a mistura toda acontecendo... Veja que agora temos restrição e o foque da busca é pelo "rombo" que levaremos. Sendo assim, a métrica desse modelo deve ser algo como *ASE*, *RMSE*, *AIC*, *BIC* ou algo do tipo, pois a proposta é chegar na melhor estimativa da probabilidade de fraude e mutiplicar esse valor pelo valor comprado conosco, para então ordenarmos pelos maiores prejus pro nosso varejo e assim a mesa foca no que causaria um maior impacto.

Que sensacional, hein?

05.

CONCLUSÃO

CONCLUSÃO



Para tudo em análise (modelagem, machine learning, inteligência artificial, seja lá como você quiser chamar isso...) é necessário estratégia! Estratégia do que é mais interessante para o problema que estou tentando resolver.

Veja que em tudo na vida existe variabilidade (Amém! Vamos orar, pois é graças a ela que tenho emprego) e uma vez que ela existe, tudo pode acontecer... O que a ciência nos orienta é buscar caminhos de minimizar esses impactos (ou maximizar quando for de nosso interesse).

No entanto, para isso funcionar na sua maior potência é MEGA importante saber gerir as métricas de qualidade de ajuste, afinal são elas que farão seu modelo mostrar o valor do mundo da ciência de dados.

Explore todas elas, treine, compare e sempre escolha seu melhor modelo não apenas vendo um número, e sim o impacto daquilo no seu negócio.

Lembre se que as métricas podem divergir (uma escolher o modelo A e a outra escolher o modelo B, por exemplo) e a escolha estará em suas mãos (que deverá estar mega alinhada com o negócio para poder tomar a melhor decisão)...

Profundo, não? Boa sorte por aí! Fui...

ISENÇÃO DE RESPONSABILIDADE

Todas as informações contidas neste material são provenientes das minhas experiências pessoais ao longo dos anos. Embora eu tenha me esforçado ao máximo, para garantir a precisão e a mais alta qualidade dessas informações e que todas as técnicas e métodos aqui ensinados sejam altamente efetivos, eu não me responsabilizo por erros (ainda mais de português) ou omissões. Você deverá utilizar e ajustar as informações deste material de acordo com sua situação e necessidades.

DIREITOS AUTORAIS

Este material está protegido por leis de direitos autorais. Todos os direitos são reservados. Você não tem permissão para vender este material nem para copiar/reproduzir o conteúdo em sites, blogs, jornais ou quaisquer outros veículos de distribuição e mídia. Qualquer tipo de violação dos direitos autorais estará sujeita a ações legais.

Conforme a Lei 9.610/98, é proibida a reprodução total e parcial ou divulgação comercial sem a autorização prévia e expressa do autor (artigo 29).

Imagens ilustrativas por freepik.

DIVIRTA-SE APRENDENDO DE VERDADE!

Sobre Adriana Silva

Graduada em Estatística pela Universidade Estadual Paulista Júlio de Mesquita Filho (Unesp – 2008), Mestre em Ciências: Estatística e Experimentação Agronômica pela Universidade de São Paulo (ESALQ/USP – 2012). Autora do livro "Data Mining Techniques to Acquire New Customers", coordenadora e docente no curso "SAS Academy Data Science" do SAS, além de professora de Estatística e Machine Learning em trabalhos voluntários.

Vice Presidente eleita do CONRE-3 em 2019 e Analytics Director na Kantar Brasil. Já atuou como Head de Analytics no SAS, foi gerente analítica na unidade de negócio ABD (Abril Big Data) na Editora Abril e pré vendas na Oracle.

Tem como lema de vida o compartilhamento do pouco que sabe. Acredita que a ciência irá mudar o mundo e que todos tem capacidade de se divertir analisando dados, uma vez que seja bem feito!

/in/adrianamms dri@asn.rocks www.asn.rocks

Métricas de qualidade de ajuste

No mundo dos dados, nunca devemos ter um modelo preferido. O ideal é sempre testar tipos de algorítimos diferentes e escolher o melhor. Este e-book irá gerar *insights* e orientá-lo na escolha da melhor métrica para definir seu modelo campeão.

Data: fevereiro/2019

Divirta-se aprendendo de verdade!

www.asn.rocks