modelo

Livia

1/23/2019

## Análise inferencial e preditiva das decisões do Supremo Tribunal Federal

Na seção anterior, realizamos a análise exploratória dos dados. A análise exploratória é útil porque, para além de revelar a estrutura subjacente dos dados, permite também identificar possíveis anomalias nos dados e oferece bons “insights” sobre a relação entre a variável resposta e as variáveis preditoras, como foi possível verificar no teste do chi-quadrado e no valor da informação (IV). No entanto, essas medidas de significância e de força de associação têm alcance limitado. Elas são basicamente importantes porque permitem corrigir inconsistências, erros de classificação, observar e ajustar anomalias (outliers, missing, skewness), selecionar as variáveis que verdadeiramente impactam o resultado e realizar o que chamamos de feature engineering, ou seja, repensar a categorização dos dados, e.g., inicialmente se manteve os órgãos colegiados (plenário, primeira turma e segunda turma) em categoria separadas, mas se notou que essas tinham efeito muito similares. Ao agrupá-las na categoria colegiadas, verificou-se que seu o poder explicativo melhorou.

A presente seção consiste na modelagem dos dados, isto é, criar uma representação simplificada da relação entre as variáveis explicativas e a variável resposta. Seu objetivo é medir o efeito de cada variável explicativa sobre a variável resposta (i.e. decisao), isolando este efeito dos demais efeitos. A variável resposta é binária, i.e., ela assume dois valores mutuamente exclusivos “improcedente” e “procedente”.

A modelagem de dados pressupõe que as observações da variável resposta estejam aleatoria e independentemente distribuídas. Outro pressuposto é o de que sua distribuição respeite uma distribuição probabilística específica, denominada distribuição de Bernoulli. Um modelo explicativo de uma determinada distribuição probabílistica é definido e caracterizado por parâmetros. Por essa razão, essas distribuições são chamadas distribuições paramétricas. Uma vez que estamos trabalhando com uma amostra das decisões, o esforço do analista é encontrar parâmetros desta distribuição amostral que se assemelhem aos parâmetros da distribuição real dos dados. A tarefa é maximizar a verossimilhança dos parâmetros encontrados a partir da modelagem na amostra em relação aos parâmetros reais da população(Hilbe 2017,Agresti (2018), Agresti (2003)).

Para estimar a probabilidade de resposta favorável à reclamação, adotamos a regressão logística binária, que é um dos procedimentos estatísticos mais utilizados em pesquisa. Ela é considerada como uma das mais importantes rotinas estatísticas em áreas como análise de cuidados de saúde, estatísticas médicas, avaliação de crédito, ecologia, estatísticas sociais e econometria(Hilbe 2017). Com o aumento das pesquisas quantitativas em direito, a tendência é que ela se popularize entre os pesquisadores, vez que explicar o comportamento judicial consubstanciado nas decisões de mérito, cautelares ou liminares, constitui uma área de particular interesse do pesquisador do direito.

Há uma boa razão para essa popularidade. Ao contrário da regressão tradicional linear, ou normal, a regressão logística é apropriada para modelar uma variável binária. Uma variável binária tem apenas dois valores - 1 e 0. Esses valores podem ser considerados como “sucesso” e “fracasso”, ou no direito, poderíamos falar em “provido” “improvido”, “procedente”, “improcedente”, “deferido” e “indeferido”, “condenado” e “absolvido”, “concedido” e “denegado”. Se uma pesquisadora modelar uma variável binária 1/0 em um ou mais preditores usando regressão linear, i.e. quando a variável resposta é quantitativa, e.g. salário, peso e renda, os pressupostos nas quais o modelo linear é baseado são violados. Ou seja, o modelo linear não é apropriado para modelar dados binários, somente para dados quantitativos(Hilbe 2017).

A regressão logística é tipicamente usada por pesquisadores e analistas em geral para três finalidades: 1. Para prever a probabilidade de que a variável resposta resposta seja igual a 1 2. Para categorizar os resultados ou previsões 3. Para acessar as probabilidades ou riscos associados aos preditores do modelo (Hilbe 2017)

A abordagem assumida aqui é a do realismo jurídico. Diferentemente do tecnicismo jurídico, o qual assume que o comportamento judicial irá se adequar às definições legais, o realismo jurídico toma em conta a prática judicial ou as possibilidades reais do comportamento judicial, independentemente de este se adequar ao que a lei ou a doutrina postulam como ideiais.

### Divisão da base em treino e teste

Uma vez que a base é relativamente grante, 6372 decisões, optou-se por excluir os casos com dados faltantes para a variável assunto, o que provocou um redução a 87 observações. Além disso, foram rodados não somente modelos estatísticos, mas também modelos de aprendizado de máquina (machine learning). Em poucas linhas, a diferença entre modelos estatísticos e modelos de aprendizado de máquina é que os primeiros partem do pressuposto de que os dados foram gerados respeitando uma equação matemática. No caso da regressão logística, os dados respeitam a distribuição de Bernoulli, isto é, segue a mesma lógica de valores obtidos quando lançamos uma moeda, que pode dar cara ou coroa.

Nos modelos de aprendizado de máquina não se supõe qualquer fórmula explicativa da distribuição dos dados, ao contrário, para eles entre a explicação e o resultado há uma caixa preta, o que importa é identificar padrões recorrentes nesses dados, por exemplo, quase sempre que uma decisão monocrática é agravada, verifica-se que o órgão colegiado confirma a decisão. As exceções à regra (padrão) são aos poucos incorporadas ao modelo até achar um explicação ótima(Breiman and others 2001).

Os modelos de aprendizado de máquina utilizados serão boosting(T. Chen et al. 2018) e floresta aleatória(Wright and Ziegler 2017). Esses dois classificadores são aperfeiçoamentos de um outro classificador, chamado árvores de decisão. Árvores de decisão são processos de classificação das observações, no presente caso as reclamações ao STF, ordenando-as num esquema similar a uma árvore invertida, ou seja, começando do tronco, mas de cima para baixo. Cada nó da árvore representa um atributo, por exemplo, órgão julgador e segmento. Floresta aleatória utiliza processo similar, mas a partir de dezenas ou centenas de árvores, tirando-se a média delas. Boosting constrói as árvores em passos, aprendendo e utilizando-se dos erros anteriores(Jesus Filho 2017).

Para rodar os modelos, separamos as bases em duas, base treino e base teste. A base treino, com 70% das observações, contêm 4716 linhas. Como a base é desequilibrada, isto é, 2721 casos com decisões improcedentes na base treino e 1995 casos procedentes na mesma base, optamos por equilibrá-la, replicando parte dos casos de improcedência, selecionados aleatoriamente para igualá-los aos casos de improcedência.

Além disso, para melhorar o desempenho, utilizamos validação cruzada 10-fold com cinco repetições, ou seja, a base treino é dividida em 10, um desses grupos é deixado de fora, roda-se o modelo nos demais nove e verifica-se a área debaixo da curva após aplicar o resultado décima parte deixada de fora. Esse procedimento é repetido para cada uma das dez subdivisões da base treino. Volta-se a repetir todos esses processos por mais quatro vezes. Isso assegura que os melhores parâmetros e hiperparâmetros foram encontrados, aumentando assim a acurácia do modelo explicativo.

Por fim, para rodar o modelo de regressão logística, optou-se por excluir a variável instância porque esta possui baixíssima associação com a variável resposta e poderia absorver o efeito de outras variáveis. Além disso, notou-se que as variáveis paradigma, segmento, novo\_cpc e assunto, sozinhas tinham baixo poder preditivo, mas quando em interação, formam uma “nova variável” com melhor poder preditivo.

Os modelos de machine learning são resistentes a violações dos pressupostos dos modelos lineares generalizados, de modo que não há necessidade de excluir qualquer variável do base e tampouco excluir os dados faltantes. Por outro lado, se eles além de resistentes, oferecem melhor desempenho em termos preditivos, sua capacidade explicativa é limitada.

## Resultados da regressão logística

Independentemente da regressão logística, a probabilidade geral de uma reclamação ser procedente pode ser facilmente verificada somando-se as frequências e verificando a relação de cada uma delas com o total:

[inserir tabela prob\_geral]

Da tabela acima, verifica-se que a probabilidade de improcedência, 0,577, é superior a de procedência, 0,423. Se dividirmos um valor pelo outro, constatamos que a improcedência é 1,36 vezes mais provável que a procedência. Essa proporção não é grande, vez que em outras áreas do direito, a taxa de improcedência é muito maior. Como perdemos alguns poucos casos de improcedência ao realizar os filtros, esse número é seguramente um pouco maior.

Para efeitos de interpretação dos coeficientes, recorde-se que o **R** atribui como categoria positiva aquela que aparece em primeiro lugar na ordem alfabética. Assim, improcedente é a classe “positiva” e procedente é a “negativa”. Isso significa que, ao ler os coeficientes, aqueles cujas razões de possibilidade (odds ratio - or) estão próximos de zero tendem a improceder, aqueles muitos distantes de zero tendem a dar procedência.

Como as variáveis são todas categóricas, o intercepto assume categoria base. Assim, se observarmos na tabela abaixo que reporta os estimadores, o nome do ministro Alexandre de Moraes não aparece, porque ele é a categoria de referência, ou seja, o intercepto, isto é -2.72938. Para calcular o coeficiente dos demais ministros, é necessário somá-los a este valor. Feito isso, é importante lembrar que o resultado vem em termos de log odds. Para saber quais as chances de uma reclamação ser procedente ou improcedente se for distribuída para um ministro ou outro, é necessário exponenciar esses coeficientes e dividir um pelo outro. Para dar um exemplo, se quisermos comparar o quanto muda a chance de procedência de um ministro para o outro, iremos comparar dois ministros que se encontram ordinariamente em polos distintos. Feito tal cálculo, chega-se ao seguinte resultado:

Uma reclamação dirigida ao Edson Fachin tem mais de 46 vezes chances de ser improcedente do que para o Alexandre de Moraes.

## Resultados para os modelos de aprendizado de máquina

Os resultados dos modelos floresta aleatória e boosting foram muito similares. A acurácia é mantida em 0,85, a sensitividade e a especificidade são mantidas igualmente em 0,85. Este é um excelente equilíbrio entre as resposta. Por essa razão, iremos reportar somente os resultados para boosting.

Diferentemente de regressão logística, não é possível apresentar coeficientes para boosting e explicar em quanto aumenta a probabilidade de concessão quando passamos de uma categoria para outra, como fizemos para regressão logística. Todavia, machine learning tem melhor poder preditivo e é possível de modo geral a importância de cada variável para o resultado. A seguir apresentamos os resultados em forma de cenários.

Uma vez que parte da base, 30%, ou seja, 1592 casos, foi reservada e mantida intocada para a validação do modelo, a presente subseção será utilizada para realização da análise preditiva. A análise preditiva é importante não só porque nos permite informar, para casos concretos, a probabilidade de uma decisão judicial da Suprema Corte conceder o pedido de reclamação, mas também porque é possível calcular métricas de desempenho, tais como verificar a capacidade de acerto do modelo. Igualmente, é possível verificar probabilidades ótimas para elevar a capacidade de predizer decisões favoráveis ou decisões desfaráveis, conforme o interesse em um ou em outro desfecho.

Apresentaremos quatro cenários distintos de reclamação a fim de ilustrar tanto a capacidade preditiva do modelo quanto para mostrar como variam as chances de procedência ou improcedência a depender do perfil do caso

[colocar aqui os quatro cenários]

### Imporância das variáveis

[Incluir importância das variáveis]

## Interpretação dos resultados

Os três modelos, regressão logística, extreme gradiente boosting e floresta aleatória são bastante similares nos resultados. Boosting e Floresta aleatória desempenham melhor em termos da acurácia, mas a regressão logística é melhor para explicar o efeito de cada uma das variáveis sobre a decisão judicial.

Os resultados mostram relevante significância estatística e prática para as órgãos julgadores e para o segmento do Judiciário. Eles sugerem que há maior chance de provimento quando as reclamações são julgadas pelos ministros Alexandre de Moraes e Gilmar Mendes. Por outro lado, ficou evidenciado que os órgãos colegiados têm papel meramente homologatório.

Ao contrário do que comumente se esperava, o CPC 2015 não implicou em maior maior número de procedências, ao contrário, a taxa de procedência das reclamações caiu após sua vigência. Essa conclusão caminha na linha contrária do que se esperava do efeito ampliador das hipóteses de cabimento da reclamação promovido pelo CPC 2015.

## Métricas de desempenho

### Curva ROC

A curva ROC é um bom instrumento para verificação dos pontos de corte (thresholds) de probabilidades para respostas binárias, ou seja, os pontos de probabilidade em que teremos melhores predições de decisões favoráveis ao passo em que controlamos o número de acertos para decisões desfavoráveis. Observando os gráficos de é possível verificar do primeiro deles que em torno de 0.45, ou um pouco menos, as taxas de verdeiros positivos (tpr ou sensitividade) passam a aumentar, chegando a algo próximo de 80%. Se reduzirmos o ponto de corte para melhorar as predições de verdadeiros positivos, teremos as taxas de falsos negativos elevadas, ou seja, erraremos muito nas predições de não provimento. O segundo gráfico é mais ilustrativo e permite identificar pontos de corte ótimos. As linhas da sensitividade e da especificidade se cruzam no ponto de corte 0,45. No entanto, se observarmos bem a linha da especificidade, notamos que entre os pontos 0,35 e 0,45 de probabilidade, seu declive é mínimo, enquanto que a linha da sensitividade mantêm-se em aclive. De modo que podemos elevar o ponto de corte para 0,43 ou 0,45 e assim melhorar os acertos para a resposta positiva, sem afetar significativamente os acertos para respostas negativas.

A curva ROC fornece uma outra métrica, a área sob a curva (area under the curve - AUC), pela qual se verifica o quão distante nosso modelo está da aleatoriedade da resposta. Ou seja, ela informa o quanto o modelo preditivo difere de uma aposta baseada no lançamento de uma moeda. Suponha uma linha diagonal saindo ponto (0,0) até o ponto (1,1) do primeiro gráfico. Esta curva teria uma área de 0.5 sob ela (o triângulo inferior à direita), indicando total aleatoriedade. Por outro lado, uma linha que saísse do ponto (0,0), passando pelo ponto (0,1) e terminasse no ponto (1,1), teria uma área de 1 (área de todo o quadrante), indicando 100% de acerto. A curva do gráfico, como mostra o valor anotado no interior do gráfico, está sobre uma área de mais de .93, o que indica um desempenho excelente do modelo. Isto é, está longe da aleatoridade.

[Colocar curva roc aqui]

Tomando a visualização da curva roc em consideração, abaixo mostramos as métricas de desempenho do modelo na base teste. A acuracia alcançou 85%, sendo que a sensitividade, ou seja, taxa de provimentos corretamente preditos, foi de 85%. Já a especificidade, taxa de improvimentos corretamente preditos, alcançou 85%. Fixamos o ponto de corte em 0,4, isto é, novos casos serão classificados como providos quando a probabilidade de provimento estiver acima de 0,4.

[Colocar aqui conf\_mat]

## Conclusão

1- O CPC/15 não elevou a taxa de provimento, ao contrário, após sua vigência as taxas de provimento cairam. 2 - Nao se pode afirmar que há um descumprimento sistemeatico das cortes inferiores, no entanto a justiça do trabalho é consistentemente recalcitrante.

3- A figura do ministro é crucial. As chances de procedência entre um ministo e outro se alteram significativametne.

A pesquisa buscou verificar duas hipóteses:

1. Se há descumprimento sistemático das decisões da Suprema Corte por parte das cortes inferiores.
2. Se o advento do CPC 2015 resultou na redução do descumprimento.

Quanto à primeira hipótese, os dados não revelam um descumprimento sistemático das decisões do STF por parte do STJ, da Justiça Eleitoral, da Justiça Federal e da Justiça Estadual. No entanto, a Justiça do Trabalho foi aquela que mais sofreu reclamações. Mais da metade das reclamações tiveram como reclamada a Justiça do Trabalho. Os resultados mostraram significância estatística para a Justiça do Trabalho.

Agresti, Alan. 2003. *Categorical Data Analysis*. Vol. 482. John Wiley & Sons.

———. 2018. *An Introduction to Categorical Data Analysis*. Wiley.

Breiman, Leo, and others. 2001. “Statistical Modeling: The Two Cultures (with Comments and a Rejoinder by the Author).” *Statistical Science* 16 (3). Institute of Mathematical Statistics: 199–231.

Chen, Tianqi, Tong He, Michael Benesty, Vadim Khotilovich, Yuan Tang, Hyunsu Cho, Kailong Chen, et al. 2018. *Xgboost: Extreme Gradient Boosting*. <https://CRAN.R-project.org/package=xgboost>.

Hilbe, Joseph M. 2017. *Logistic Regression Models*. CRC press.

Jesus Filho, José de. 2017. “Administração Penitenciária: O Controle Da População Carcerária a Partir Da Gestão Partilhada Entre Diretores, Judiciário E Facções.” PhD thesis.

Wright, Marvin N., and Andreas Ziegler. 2017. “ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R.” *Journal of Statistical Software* 77 (1): 1–17. doi:[10.18637/jss.v077.i01](https://doi.org/10.18637/jss.v077.i01).