

Análise de Modelos de Regressão e Classificação em Inteligência Artificial

Levi Natã Monteiro Maciel
Inteligência Artificial Computacional (15-M24EF)
Universidade de Fortaleza (UNIFOR)
Fortaleza, Brasil
lnatamm@edu.unifor.br

João Victor Leles Cordeiro
Inteligência Artificial Computacional (15-M24EF)
Universidade de Fortaleza (UNIFOR)
Fortaleza, Brasil
jvleles@edu.unifor.br

Abstract—Este estudo apresenta o desenvolvimento e a avaliação de modelos preditivos supervisionados em tarefas de regressão e classificação. Na primeira etapa, são explorados modelos de regressão para estimar quantitativamente a potência gerada com base na velocidade do vento, empregando abordagens como Mínimos Quadrados Ordinários (MQO) e regularização de Tikhonov. Na segunda etapa, são utilizados classificadores gaussianos para identificar expressões faciais por meio de sinais de eletromiografia. A validação dos modelos é realizada por meio de simulações de Monte Carlo, permitindo a comparação de seu desempenho em termos de acurácia.

Index Terms—Regressão, Classificação, Mínimos Quadrados Ordinários, Regularização Tikhonov, Classificadores Gaussianos, Monte Carlo, Acurácia, Inteligência Artificial.

I. INTRODUÇÃO

A modelagem preditiva em inteligência artificial é amplamente empregada para resolver problemas de previsão e classificação em diversas áreas. No contexto da previsão quantitativa, a regressão desempenha um papel essencial ao estabelecer a relação entre variáveis dependentes e independentes. Um dos métodos mais utilizados para esse fim é o de Mínimos Quadrados Ordinários (MQO), conhecido por sua simplicidade e eficiência. No entanto, em cenários onde os dados apresentam maior complexidade, pode ocorrer sobreajuste. Para mitigar esse problema, são adotadas técnicas de regularização, como a abordagem de Tikhonov, que adiciona um termo de penalização ajustado por um hiperparâmetro. λ .

Além disso, a classificação é fundamental em tarefas que buscam categorizar observações. Neste estudo, foram utilizados classificadores gaussianos para reconhecer padrões em sinais de eletromiografia captados dos músculos faciais, contribuindo para a identificação de expressões faciais. A avaliação dos classificadores considerou modelos com diferentes níveis de complexidade, incluindo versões mais sofisticadas que incorporam matrizes de covariância ajustadas por ponderação e regularização.

II. METODOLOGIA

Para cada tarefa, foram realizadas simulações de Monte Carlo com 500 iterações, utilizando uma divisão dos dados em 80% para treinamento e 20% para teste. A seguir, apresentamos as principais abordagens de modelagem e as equações fundamentais.

A. Tarefa de Regressão

Os modelos de regressão visam prever o nível de atividade enzimática com base na temperatura e pH da solução. Foram implementados os seguintes modelos:

- **MQO Tradicional:** Este modelo calcula os coeficientes β minimizando a soma dos quadrados dos erros entre as previsões e os valores reais. A fórmula para o cálculo dos coeficientes é dada por:

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

onde X é a matriz de variáveis independentes (incluindo o termo de intercepto) e y é o vetor de valores dependentes (potência gerada).

- **MQO Regularizado (Tikhonov):** Este modelo incorpora um termo de regularização para controlar o sobreajuste. A fórmula para os coeficientes é:

$$W = (X^T X + \lambda I)^{-1} X^T Y$$

onde λ é o hiperparâmetro que regula o peso do termo de penalização. Foram testados valores de λ de 0 a 1 em incrementos de 0,25.

- **Média dos Valores Observados:** Um modelo de referência que utiliza a média dos valores observados para prever y , ignorando as variáveis independentes.

B. Tarefa de Classificação

Para a tarefa de classificação, utilizamos sinais de eletromiografia capturados de músculos faciais. Os modelos gaussianos foram aplicados da seguinte forma:

- **MQO Tradicional:** Realiza a previsão de classes com base em uma regressão direta, onde a classe é determinada pela proximidade com a média mais próxima.
- **Classificador Gaussiano Tradicional:** Calcula uma matriz de covariância específica para cada classe e atribui a nova amostra à classe com a maior probabilidade condicional. A probabilidade é calculada utilizando o critério discriminante quadrático:

$$g_i(x_n) = -\frac{1}{2} \ln |\Sigma_i| - \frac{1}{2} (x_n - \mu_i)^T \Sigma_i^{-1} (x_n - \mu_i)$$

- **Classificador Gaussiano com Covariâncias Iguais:** Utiliza uma única matriz de covariância para todas as classes,

reduzindo a complexidade e aumentando a estabilidade do classificador.

- **Classificador Gaussiano com Matriz Agregada:** Baseado em uma matriz de covariância agregada ponderada por cada classe, calculada como:

$$\Sigma_{agregada} = \sum_{i=1}^C P(y_i) \Sigma_i$$

- **Classificador Gaussiano Regularizado (Friedman):** Utiliza uma combinação ponderada entre as matrizes de covariância individuais e a matriz agregada. A fórmula para a matriz de covariância regularizada é:

$$\Sigma_i^\lambda = \frac{(1 - \lambda)(n_i \cdot \Sigma_i) + (\lambda \cdot N \cdot \Sigma_{agregada})}{(1 - \lambda)n_i + \lambda \cdot N}$$

com valores de λ de 0 a 1, em incrementos de 0,25. Para $\lambda = 1$, utilizamos seguinte função discriminante:

$$g_i(x_n) = (x_n - \mu_i)^T \Sigma^{-1} (x_n - \mu_i).$$

- **Classificador de Bayes Ingênuo:** Assume independência entre os atributos, com a matriz de covariância sendo diagonal. Isso simplifica o cálculo da inversão de matriz e melhora o desempenho em alta dimensionalidade.

III. RESULTADOS

As tabelas I e II apresentam os resultados das simulações para regressão e classificação, respectivamente, incluindo a média, desvio-padrão, maior e menor valor de RSS (regressão) e acurácia (classificação).

A. Resultados da Tarefa de Regressão

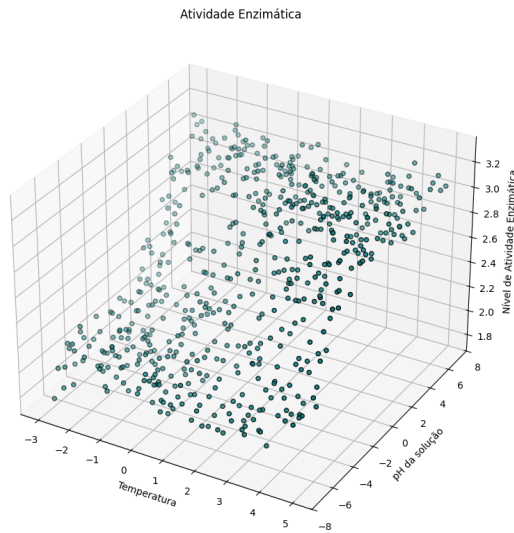


Fig. 1. Gráfico ilustrativo dos resultados da tarefa de regressão

TABLE I
RESULTADOS DA TAREFA DE REGRESSÃO

Modelo	Média	Desvio-Padrão	Maior Valor	Menor Valor
MQO Tradicional ($\lambda = 0$)	4.3215	0.4127	5.7228	3.3923
MQO Regularizado ($\lambda = 0.25$)	4.3219	0.4128	5.7259	3.3898
MQO Regularizado ($\lambda = 0.5$)	4.3229	0.4130	5.7294	3.3878
MQO Regularizado ($\lambda = 0.75$)	4.3243	0.4133	5.7334	3.3861
MQO Regularizado ($\lambda = 1$)	4.3262	0.4136	5.7378	3.3848
Media da variável dependente	22.7534	1.2228	25.8904	18.8603

B. Resultados da Tarefa de Classificação

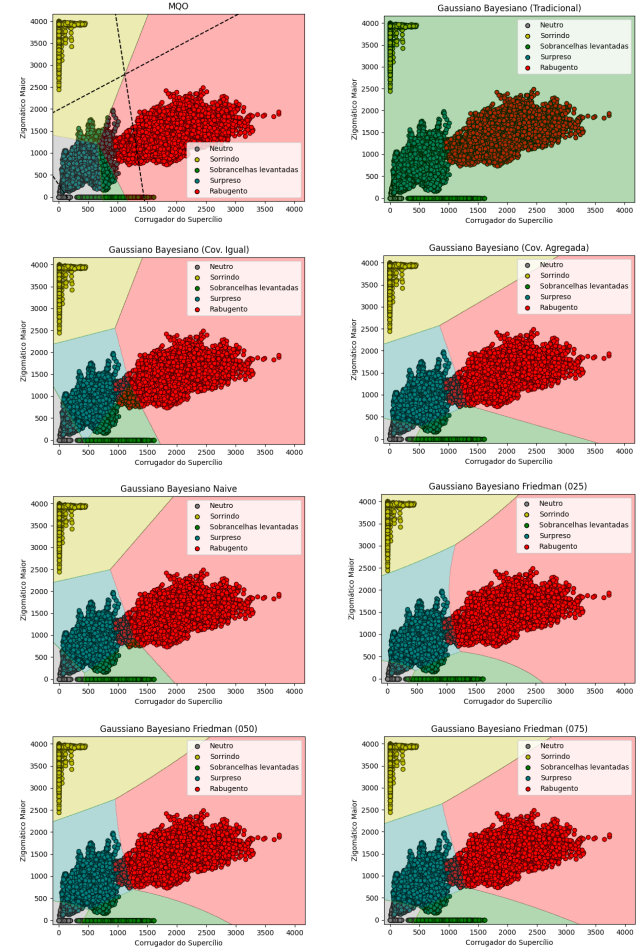


Fig. 2. Gráficos ilustrativos dos limites de decisão de cada modelo.

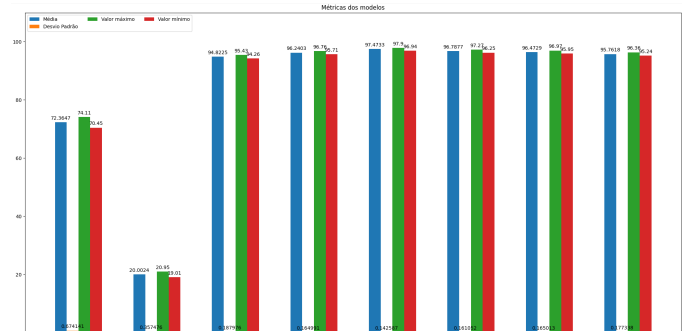


Fig. 3. Gráfico ilustrativo das métricas dos modelos de classificação.

TABLE II
RESULTADOS DA TAREFA DE CLASSIFICAÇÃO

Modelos	Média da Acurácia	Desvio Padrão	Maior Acurácia (%)	Menor Acurácia (%)
MQO Tradicional	72.36%	0.67	74.11%	70.45%
Classificador Gaussiano Tradicional	20.00%	0.35	20.95%	19.00%
Classificador Gaussiano (Cov. Treino)	94.82%	0.18	95.43%	94.26%
Classificador Gaussiano (Cov. Agregada)	96.24%	0.16	96.76%	95.71%
Classificador de Bayes Ingênuo	95.76%	0.17	96.36%	95.24%
Class. Gauss. Regularizado (Friedman $\lambda = 0.25$)	97.47%	0.14	97.89%	96.94%
Class. Gauss. Regularizado (Friedman $\lambda = 0.5$)	96.78%	0.16	97.27%	96.25%
Class. Gauss. Regularizado (Friedman $\lambda = 0.75$)	96.47%	0.16	96.97%	95.95%

IV. CONCLUSÕES

A análise comparativa dos modelos de regressão demonstrou a influência da regularização nos resultados obtidos. O modelo tradicional Mínimos Quadrados Ordinários (MQO) produziu uma média de 4,3215 e um desvio padrão de 0,4127, enquanto os modelos regularizados exibiram pequenas flutuações nesses valores com o aumento do parâmetro λ . Ao ajustar λ , observou-se uma tendência ligeiramente crescente na média e no desvio padrão, sugerindo que a regularização introduz controle adicional sobre a variabilidade do modelo. Além disso, os valores máximo e mínimo seguiram uma tendência semelhante, indicando uma influência moderada da regularização na amplitude dos resultados. Comparado à média da variável dependente (22,7534), é evidente que os modelos de regressão operam em uma escala significativamente menor, reforçando a necessidade de ajustes apropriados para garantir previsões robustas.

Na tarefa de classificação, os classificadores gaussianos com matrizes de covariância regularizadas demonstraram os melhores desempenhos em termos de acurácia. O Classificador Gaussiano Regularizado (Friedman) com $\lambda = 0.25$ obteve a maior média de acurácia (97.47%), seguido pelo modelo com matriz de covariância agregada (96.24%) e pelo classificador de Bayes Ingênuo (95.76%). Esses resultados indicam que a regularização adequada da matriz de covariância contribui para um melhor equilíbrio entre viés e variância, reduzindo a sensibilidade a variações nos dados.

Por outro lado, o Classificador Gaussiano Tradicional apresentou um desempenho significativamente inferior (20.00%), evidenciando a importância de técnicas de ajuste para lidar com distribuições complexas. O MQO Tradicional, apesar de alcançar 72.36% de acurácia, mostrou maior variabilidade nos resultados, o que pode impactar sua confiabilidade em aplicações práticas.

Esses achados reforçam que, ao trabalhar com dados de alta dimensionalidade e características heterogêneas, a escolha adequada dos parâmetros de regularização é fundamental para garantir um desempenho preditivo robusto e estável.

REFERENCES

- [1] Msc. Prof. Paulo Cirillo Souza Barbosa, "Inteligência Artificial Computacional - T296 (Slides dos blocos 5 e 6)," Centro de Ciências Tecnológicas - CCT, Fortaleza, Ceará, Brasil.