

Implementação e Avaliação de Modelos de Inteligência Artificial em Energia Eólica e Expressões Faciais

Caio Barros da Costa - 2315082, João Pedro Ribeiro Mendes - 2315069

Ciências da Computação - Unifor
Fortaleza, Brasil

Abstract—Este artigo investiga, por meio de implementações em alto nível utilizando *NumPy* e *Matplotlib*, a aplicação de modelos clássicos de aprendizagem supervisionada. No cenário de regressão, foi usada uma base de dados de um aerogerador com Mínimos Quadrados Ordinários (MQO) e sua variação regularizada (Ridge). Já na tarefa de classificação, exploramos sinais mioelétricos coletados pelo sensor Myoware Muscle Sensor em conjunto com um microcontrolador ESP32, discriminando cinco categorias de expressões faciais: Neutro, Sorriso, Sobrancelhas Levantadas, Supreso e Rabugento. Foram implementados e comparados com MQO Tradicional, Classificador Gaussiano Tradicional, Classificador Gaussiano com Covariâncias Iguais, Classificador Gaussiano com Matriz Agregada, Classificador Gaussiano Regularizado (Friedman), Classificador de Bayes Ingênuo. O desempenho foi avaliado em termos de métricas estatísticas sobre repetições de Monte Carlo, evidenciando diferenças estruturais entre hipóteses de covariância e efeitos da regularização.

Index Terms—NumPy, Matplotlib, MQO, Ridge, Classificação Gaussiana, Naive Bayes, EMG, Aerogerador.

I. INTRODUÇÃO

Este trabalho avalia modelos clássicos de aprendizagem supervisionada implementados diretamente com *NumPy* e visualizados no *Matplotlib*. Foram considerados dois cenários: (i) regressão univariada em dados de um aerogerador, com MQO e Ridge; e (ii) classificação multiclasse de sinais mioelétricos obtidos pelo sensor Myoware e um ESP32, abrangendo MQO Tradicional, Classificadores Gaussianos (em diferentes hipóteses de covariância e regularização) e Bayes Ingênuo. A análise, baseada em validação Monte Carlo, compara o impacto das hipóteses probabilísticas e da regularização no desempenho dos modelos.

II. MATERIAIS E MÉTODOS

A. Bases de Dados

Aerogerador (Regressão). Utilizou-se a relação entre velocidade do vento e potência gerada. **EMG (Classificação).** Os sinais foram adquiridos com o sensor Myoware Muscle Sensor acoplado a um ESP32, permitindo identificar expressões faciais em cinco classes: Neutro, Sorriso, Sobrancelhas Levantadas, Surpreso e Rabugento.

B. Particionamento e Protocolo

A avaliação seguiu o esquema de validação Monte Carlo com 500 repetições, utilizando divisões aleatórias em 80%

treino e 20% teste. Para cada repetição foram calculadas métricas de desempenho, de onde extraímos média, desvio-padrão, maior e menor valores.

C. Modelos de Regressão

Foram implementados três modelos de regressão: a média dos valores observados (baseline), o MQO tradicional e o MQO regularizado na forma de Tikhonov (Ridge). Em todos os casos, o intercepto também é estimado juntamente com os coeficientes.

Para cada um dos modelos, nos baseamos nas seguintes formas matemáticas:

MQO Tradicional:

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

MQO Regularizado:

$$\hat{\beta}_{ridge} = (X^T X + \lambda I)^{-1} X^T y \quad (2)$$

D. Modelos de Classificação

Na etapa de classificação, foram implementados os seguintes modelos: MQO tradicional, Classificador Gaussiano Tradicional, Classificador Gaussiano com Covariâncias Iguais, Classificador Gaussiano com Matriz Agregada, Classificador Gaussiano Regularizado (Friedman) e o Classificador de Bayes Ingênuo.

Para cada um dos modelos, nos baseamos nas seguintes formas matemáticas:

MQO Tradicional:

$$\hat{W} = (X^T X)^{-1} X^T Y \quad (3)$$

Classificador Gaussiano Tradicional

$$g_i(x) = \log \pi_c - \frac{1}{2} \log |\Sigma_c| - \frac{1}{2} (x - \mu_c)^T \Sigma_c^{-1} (x - \mu_c) \quad (4)$$

Classificador Gaussiano com Covariâncias Iguais:

$$\delta_c(x) = x^T \Sigma^{-1} \mu_c - \frac{1}{2} \mu_c^T \Sigma^{-1} \mu_c + \log \pi_c \quad (5)$$

Classificador Gaussiano com Matriz Agregada:

$$\Sigma = \frac{1}{N - C} \sum_{c=1}^C (n_c - 1) \Sigma_c \quad (6)$$

Classificador Gaussiano Regularizado (Friedman):

$$\Sigma_c^{(\lambda)} = (1 - \lambda)\Sigma_c + \lambda\Sigma, \quad \lambda \in [0, 1] \quad (7)$$

Classificador de Bayes Ingênuo (Naive Bayes Gaussiano):

$$p(x_j | y = c) = \frac{1}{\sqrt{2\pi\sigma_{cj}^2}} \exp\left(-\frac{(x_j - \mu_{cj})^2}{2\sigma_{cj}^2}\right) \quad (8)$$

III. RESULTADOS

A. Regressão (Aerogerador)

Os resultados indicam que o modelo da média não captura a relação entre velocidade do vento e potência, apresentando elevado erro ($MSE \approx 11173.94$, $R^2 \approx -0.003$). O MQO sem intercepto melhora o ajuste, mas permanece limitado pela restrição da reta à origem ($MSE \approx 2569.95$, $R^2 \approx 0.769$). Com intercepto, o MQO alcança desempenho significativamente superior ($MSE \approx 809.46$, $R^2 \approx 0.928$).

A regressão Ridge, para diferentes valores de λ , produziu resultados praticamente idênticos ao MQO com intercepto, indicando ausência de ganhos relevantes com a regularização. Apenas para $\lambda = 100$ observou-se leve degradação do ajuste. Assim, conclui-se que o MQO tradicional é suficiente para este conjunto de dados, enquanto a regularização não exerce impacto expressivo no desempenho.

TABLE I
RESULTADOS MÉDIOS DOS MODELOS DE REGRESSÃO

Modelo	MSE	R^2
Média da variável dependente	11171.9620	-0.0028
MQO sem intercepto	2558.0139	0.7701
MQO com intercepto	807.5964	0.9276
Ridge ($\lambda = 0$)	807.5964	0.9276
Ridge ($\lambda = 0.5$)	807.5852	0.9276
Ridge ($\lambda = 1$)	807.5741	0.9276
Ridge ($\lambda = 10$)	807.4016	0.9276
Ridge ($\lambda = 100$)	808.3850	0.9275

TABLE II
DESEMPENHO EM REGRESSÃO (MSE): MÉDIA, DESVIO-PADRÃO, MÁXIMO E MÍNIMO (500 REPETIÇÕES).

Modelo	Média	Desv.	Máx.	Mín.
Média da variável dep.	11111.62	627.29	13108.52	9499.78
MQO sem intercepto	2551.26	134.88	2952.64	2179.80
MQO com intercepto	791.30	172.97	1516.64	420.37
Ridge ($\lambda = 0$)	791.30	172.97	1516.64	420.37
Ridge ($\lambda = 0.5$)	791.29	172.90	1516.29	420.47
Ridge ($\lambda = 1$)	791.29	172.84	1515.93	420.56
Ridge ($\lambda = 10$)	791.26	171.66	1509.56	422.35
Ridge ($\lambda = 100$)	793.62	160.63	1451.80	442.31

B. Classificação (EMG)

A Tabela V apresenta as acurácias obtidas pelos diferentes modelos avaliados na tarefa de classificação dos sinais EMG. Observa-se que o MQO apresentou desempenho inferior, enquanto os classificadores Gaussianos alcançaram acurácias elevadas, em especial o QDA, que obteve o melhor resultado médio. O Naive Bayes também apresentou desempenho

TABLE III
DESEMPENHO EM REGRESSÃO (SSE): MÉDIA, DESVIO-PADRÃO, MÁXIMO E MÍNIMO (500 REPETIÇÕES).

Modelo	Média	Desv.	Máx.	Mín.
Média da var. dep.	5000229.85	282279.54	5898835.00	4274899.22
MQO sem intercepto	1148066.56	60697.52	1328689.38	980090.33
MQO com intercepto	356083.82	77835.79	682489.48	189165.53
Ridge ($\lambda = 0$)	356083.82	77835.79	682489.48	189165.53
Ridge ($\lambda = 0.5$)	356082.40	77806.17	682328.71	189209.49
Ridge ($\lambda = 1$)	356081.05	77776.56	682168.09	189253.51
Ridge ($\lambda = 10$)	356068.54	77246.88	679303.30	190055.42
Ridge ($\lambda = 100$)	357129.23	72281.96	653311.62	199038.91

TABLE IV
DESEMPENHO EM REGRESSÃO (R^2): MÉDIA, DESVIO-PADRÃO, MÁXIMO E MÍNIMO (500 REPETIÇÕES).

Modelo	Média	Desv.	Máx.	Mín.
Média da var. dep.	-0.00	0.00	-0.00	-0.03
MQO sem intercepto	0.77	0.01	0.80	0.73
MQO com intercepto	0.93	0.01	0.96	0.88
Ridge ($\lambda = 0$)	0.93	0.01	0.96	0.88
Ridge ($\lambda = 0.5$)	0.93	0.01	0.96	0.88
Ridge ($\lambda = 1$)	0.93	0.01	0.96	0.88
Ridge ($\lambda = 10$)	0.93	0.01	0.96	0.88
Ridge ($\lambda = 100$)	0.93	0.01	0.96	0.89

competitivo, próximo ao dos classificadores Gaussianos. Já a versão regularizada mostrou estabilidade para valores intermediários de λ , mas desempenho significativamente reduzido para $\lambda = 1$.

TABLE V
ACURÁCIA DOS MODELOS DE CLASSIFICAÇÃO: MÉDIA, DESVIO-PADRÃO, MÁXIMO E MÍNIMO (500 REPETIÇÕES).

Modelo	Média	Desv.	Máx.	Mín.
MQO Tradicional	0.724	0.006	0.741	0.701
Gaussiano Tradicional (QDA)	0.967	0.002	0.971	0.961
Gaussiano Cov. Iguais (LDA)	0.962	0.002	0.969	0.958
Gaussiano Cov. Agregada	0.962	0.002	0.969	0.958
Naive Bayes Gaussiano	0.966	0.002	0.969	0.960
Gaussiano Reg. ($\lambda = 0$)	0.962	0.002	0.969	0.958
Gaussiano Reg. ($\lambda = 0.25$)	0.962	0.002	0.969	0.958
Gaussiano Reg. ($\lambda = 0.5$)	0.962	0.002	0.969	0.958
Gaussiano Reg. ($\lambda = 0.75$)	0.962	0.002	0.969	0.958
Gaussiano Reg. ($\lambda = 1$)	0.290	0.005	0.306	0.270

IV. DISCUSSÃO

Na análise de regressão, verificou-se que o modelo da média foi insuficiente para capturar a relação entre velocidade do vento e potência gerada. O MQO sem intercepto melhorou o ajuste, mas permaneceu limitado por impor a passagem da reta pela origem. A inclusão do intercepto resultou em ganho expressivo de desempenho, atingindo valores de R^2 próximos a 0.93. A regressão Ridge, por sua vez, apresentou métricas praticamente idênticas ao MQO com intercepto, indicando que a regularização não trouxe benefícios significativos neste conjunto de dados, uma vez que não havia problemas de instabilidade ou multicolinearidade relevantes.

No contexto de classificação, os resultados demonstraram clara superioridade dos classificadores Gaussianos em relação

ao MQO, que obteve acurácia de apenas 72.4%. O QDA apresentou o melhor desempenho (96.7%), seguido de perto pelo Naive Bayes e pelos modelos com covariância igual ou agregada, que se mantiveram acima de 96%. A regularização de Friedman mostrou-se estável para valores intermediários de λ , mas degradou drasticamente em $\lambda = 1$, sugerindo que a interpolação excessiva entre covariâncias pode comprometer a separabilidade entre classes. Esses achados ressaltam que a escolha do modelo adequado depende tanto da estrutura dos dados quanto das hipóteses probabilísticas adotadas.

V. CONCLUSÃO

Os experimentos realizados evidenciaram a efetividade dos modelos lineares e probabilísticos implementados em *NumPy* e visualizados com *Matplotlib*. No problema de regressão com dados de aerogerador, observou-se que a inclusão do intercepto no MQO foi determinante para a qualidade do ajuste, alcançando $R^2 \approx 0.93$ e reduzindo significativamente o erro em relação ao modelo da média e ao MQO sem intercepto. A regressão Ridge apresentou resultados praticamente idênticos ao MQO tradicional, confirmando que, neste conjunto de dados, a regularização não trouxe ganhos expressivos de desempenho.

Na tarefa de classificação de sinais mioelétricos, os classificadores Gaussianos obtiveram as maiores acurácias, com destaque para o QDA (96.7%) e o Naive Bayes (96.5%), ambos superando o MQO. O LDA e as variantes com matriz agregada também apresentaram desempenho consistente, enquanto a regularização de Friedman manteve estabilidade em valores intermediários de λ , mas degradou para $\lambda = 1$. Assim, conclui-se que os modelos Gaussianos são mais adequados para este tipo de dado, enquanto em regressão linear o MQO tradicional é suficiente. Esses resultados demonstram a versatilidade e a robustez de implementações diretas em *NumPy*, mesmo em contextos distintos como dados industriais e biomédicos.