Método dos Mínimos Quadrados aplicado no ajuste dos pesos de notas para disciplinas acadêmicas

A. B. C. Sobrenome
Universidade Universidade
Departamento de Departamento
Caixa Postal: XXXXX - CEP: XXXXX-XXX - Cidade - Estado
Fone: +XX-XX-XXXX-XXXX - Fax: +XX-XX-XXXX-XXXX
email@de.contato

Resumo

Estuda-se a viabilidade do uso do Método de Mínimos Quadrados para o ajuste dos pesos referentes ao cálculo de notas de uma disciplina acadêmica, de modo a maximizar o desempenho dos estudantes. São analisadas duas configurações principais relativas ao número de variáveis livres (pesos) na composição da nota, mostrando a aplicabilidade do método, em especial nos casos em que há poucas variáveis, além de como *outliers* afetam as grandezas de análise de desempenho geral (média da turma e distribuição de alunos aprovados/em exame final/reprovados) para cada uma das configurações.

Palavras-chave: Ajuste de pesos, Barreira de domínio, Cálculo de notas, Maximização de desempenho estudantil, Método dos Mínimos Quadrados.

1 Objetivo

Nesse trabalho, por meio do Método de Mínimos Quadrados, determinaram-se os pesos para o ajuste das notas de Física Básica Experimental 2, de modo a maximizar, na média, o desempenho dos estudantes.

2 Fundamentação Teórica

As notas da matéria se dividem em relatórios e provas. Até o momento de escrita desse trabalho, dispunhamos apenas das notas de dois dos três relatórios e provas realizados para o semestre, contudo, estabelecida a validade e a praticidade do seguinte método para tal aplicação como forma de *proof of concept* pode-se então aplicá-lo, mantida a infraestrutura computacional, de maneira ágil para a coleção completa das notas do semestre.

Foram estudados duas configurações de pesos, em uma delas, tomamos todos os pesos de cada uma das avaliações como independentes, em outra, consideramos apenas o peso da média aritméticas dos relatórios e da média aritmética das provas. A análise da fundamentação matemática se dará para esse caso, em virtude de sua simplificada descrição, em termos do número de símbolos empregados. A análise para os pesos independentes, pode ser feita de forma completamente análoga, e sua formulação final pode ser vista na Equação 3.

Considerando as notas P e R, relativas à média artimética das provas e dos relatórios, respectivamente, e seus pesos correspondentes P_P e P_R, a média final pode ser calculada por uma média ponderada, da forma

$$N_F = \frac{P_P P + P_R R}{P_P + P_R} \tag{1}$$

e, deseja-se tentar aproximar, para todos os alunos, a nota N_F de 10. Isto é, haverá um sistema com linhas do tipo

$$P_P P + P_R R = 10 \cdot (P_P + P_R)$$

combinando fatores múltiplos do mesmo peso, é obtido

$$P_P(P - 10) + P_R(R - 10) = 0$$

onde desejam-se ajustar os pesos P_P e P_R de forma a aproximar, para todos os alunos, essa igualdade ao máximo. Entretanto, esse formato tem um problema: A solução trivial é sempre solução, portanto o método de mínimos quadrados sempre produziria, para o vetor de resultados, o vetor nulo.

Identificou-se como a origem do problema a inserção de um grau de liberdade que não está realmente presente, isto é, pode-se reduzir o problema estabelecendo uma restrição em um dos pesos, como, por exemplo, o peso P_{R2}, de forma que a soma de todos os pesos seja sempre 1. Isso permite reescrever a Equação 2 como

$$N_F = P_P P + (1 - P_P)R$$

e, ao aproximar a nota final N_F ao máximo para todos os indivíduos, resolve-se um sistema do tipo

$$P_N(P - R) = 10 - R. \tag{2}$$

O caso com os 4 pesos independentes também apresenta, ao tentar ajustar todos os pesos, a solução trivial. Novamente, se emprega uma solução análoga, fixando-se a soma dos pesos como unitária, permitindo reescrever o problema de mínimos quadrados como uma aproximação, para todos os alunos, de igualdades da forma se torna

$$P_{R1}(R_1 - P_2) + P_{R2}(R_2 - P_2) + P_{P1}(P_1 - P_2) = 10 - P_2 \tag{3}$$

onde $R1$, $R2$, P_{R1} P_{R2} são as notas e os pesos dos relatórios 1 e 2, respectivamente, e P_1 e P_{P1} são a nota e o peso da primeira prova. P_2 é a nota da segunda prova.

3 Procedimento

Estabelecida a equação a ser aproximada de maneira geral para todos os alunos, exprimiu-a em termos de um sistema linear, da forma, para o caso da Equação 2

$$\begin{bmatrix} (P^1 - R^1) \\ \vdots \\ (P^n - R^N) \end{bmatrix} \begin{bmatrix} P_P \end{bmatrix} = \begin{bmatrix} (10 - R^1) \\ \vdots \\ (10 - R^N) \end{bmatrix} \tag{4}$$

em que os sobrescritos são índices referindo aos alunos $1, 2, \cdots, N$. No caso da Equação 3, o sistema linear toma uma forma mais complexa, porém similar

$$\begin{bmatrix} (P_1^1 - P_2^1) & (R_1^1 - P_2^1) & (R_2^1 - P_2^1) \\ & \vdots & \\ (P_1^N - P_2^N) & (R_1^N - P_2^N) & (R_2^N - P_2^N) \end{bmatrix} \begin{bmatrix} P_{P1} \\ P_{R1} \\ P_{R2} \end{bmatrix} = \begin{bmatrix} (10 - R^1) \\ \vdots \\ (10 - R^N) \end{bmatrix}. \tag{5}$$

Um fator que pode ser levado em conta nessas condições envolveu a exclusão de algumas entradas com base em algum fator discriminante, refazendo então o método de mínimos quadrados. Ilustrativamente, foram calculadas as configurações de peso, em ambos os modelos, levando em conta um limiar mínimo variável para a média aritmética das notas das provas. Isso foi feito pois possibilita um estudo da forma em que alunos com menor envolvimento na disciplina (o parâmetro para medir o envolvimento sendo, nesse caso, a média das provas) afetam os resultados do processo.

De maneira direta, o procedimento pode ser subdividido em quatro etapas

1. Calcular a configuração de maximização dos pesos;
2. Incrementar um parâmetro limiar para consideração das notas;
3. Excluir as notas referentes aos alunos cuja média aritmética das provas esteja abaixo desse limiar;
4. Se restaram alunos suficientes, retornar ao item 1.

Esse procedimento foi implementado computacionalmente utilizando decomposição QR por rotações em um código na linguagem C++.

Assim, para cada um dos valores limiares, e para ambas as configurações, além dos próprios pesos foram também calculadas a média final da turma e a distribuição de alunos aprovados/em exame final/reprovados, de forma que os pesos fossem traduzidos para grandezas mais tangíveis.

É evidente também que a resolução dos problemas de mínimos quadrados capturados nas Equações 2 e 3 não impõe qualquer limitação a respeito dos valores finais que os pesos podem assumir e, por isso, surgem pesos negativos e pesos maiores do que 1, fazendo com que alunos possam ter médias finais maiores do que 10 ou ainda que um aluno com notas maiores tenha uma média final menor do que outro aluno que, contraditoriamente, teve, em todas as avaliações, notas menores.

No caso da configuração em que existem apenas dois pesos, esse problema pode ser enfrentado por um passo extra no cálculo dos dados que foi denominado como **barreira de domínio**, já que a Equação 2 pode ser enxergada como uma função que, em condições ideais, recebe valores do intervalo fechado $[0, 1]$, os pesos, e produz uma média final entre 0 e 10. Quando esses pesos, contudo, estavam fora desse domínio de validade, a barreira de domínio os levava para o valor válido mais próximos, isto é

- Se o peso está acima de 1, torne-o 1;
- Se o peso está abaixo de 0, torne-o 0;
- Caso nenhum dos dois aconteça, não modifique seu valor.

Note que esse dispositivo tem aplicação mais dificultada no modelo de pesos independentes, pois é possível que, embora um peso seja negativo, os demais ainda estejam no intervalo $[0, 1]$. Portanto, a análise com barreira de domínio se limitou ao modelo de pesos para as médias.

4 Resultados e Discussões

Utilizando o procedimento delineado, para as o procedimento de pesos na primeira configuração, obtiveram-se os dados que constam na Tabela 1.

Nessa tabela, é interessante notar o comportamento intuitivo de que, conforme o limiar de notas, denotado como "Nota de corte" se torna mais alto, isto é, mantém-se para a maximização apenas os alunos com maiores notas na prova, o peso do relatório se torna cada vez menor. Além disso, ao considerar a turma como um todo, pela forma como o peso da prova é negativo na primeira linha, tem-se uma base para afirmar que os alunos obtiveram um desempenho melhor nos relatórios do que nas provas.

Apartir da Tabela 1, também foram feitos dois gráficos para uma apreensão mais visual do comportamento dos pesos *versus* a nota de corte. Eles podem ser vistos nas Figuras 1 e 2.

No caso dos pesos independentes, os valores encontrados para os pesos em função da nota de corte estão na Tabela 2 e no gráfico da Figura 3.

É importante notar que, embora no modelo de pesos para as médias dos relatórios e das provas existam três notas de corte para as quais são encontrados unicamente valores válidos, não há nenhuma nota de corte para a qual isso ocorra no modelo de pesos independentes. Isso

	Pesos com BD		Pesos sem BD	
Nota de corte	Provas	Relatórios	Provas	Relatórios
0.0	0.000000	1.000000	-0.454422	1.4544200
0.2	0.000000	1.000000	-0.260823	1.2608200
1.1	0.000000	1.000000	-0.052136	1.0521400
3.0	0.179256	0.820744	0.179256	0.8207440
3.2	0.695018	0.304982	0.695018	0.3049820
6.1	0.897815	0.102185	0.897815	0.1021850
6.3	1.000000	0.000000	1.081570	-0.0815747
7.1	1.000000	0.000000	1.353800	-0.3537990
8.5	1.000000	0.000000	1.325910	-0.3259050
8.8	1.000000	0.000000	1.180080	-0.1800770
9.3	1.000000	0.000000	1.111110	-0.1111110
9.6	1.000000	0.000000	1.000000	0.0000000

Tabela 1: Tabela de pesos obtidos pela configuração que leva em conta as médias aritméticas dos relatórios e das provas, para os casos com e sem barreira de domínio. As configurações nas quais ambos os pesos são positivos (e portanto, não são afetados pela barreira de dominio) estão destacados em cinza. A representação em gráfico dessa tabela pode ser visto nas Figuras 1 e 2.

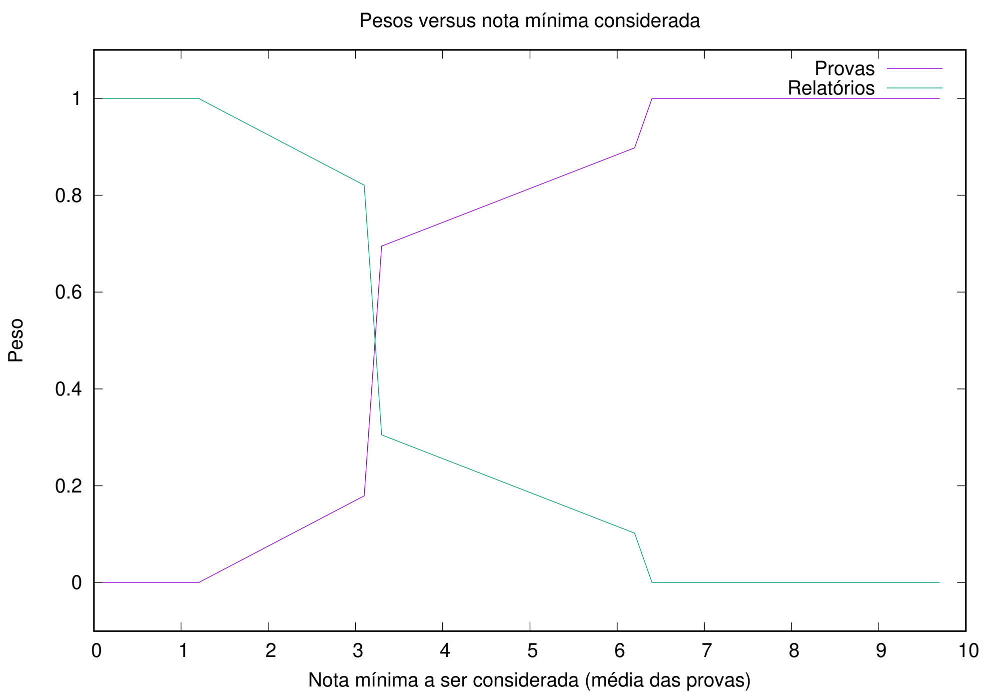

Figura 1: Pesos para a média das provas (em roxo) e dos relatórios (em verde), em função da nota mínima considerada para o método de mínimos quadrados com a barreira de domínio. Os valores numéricos dos pontos presentes nesse gráfico podem ser vistos na Tabela 1.

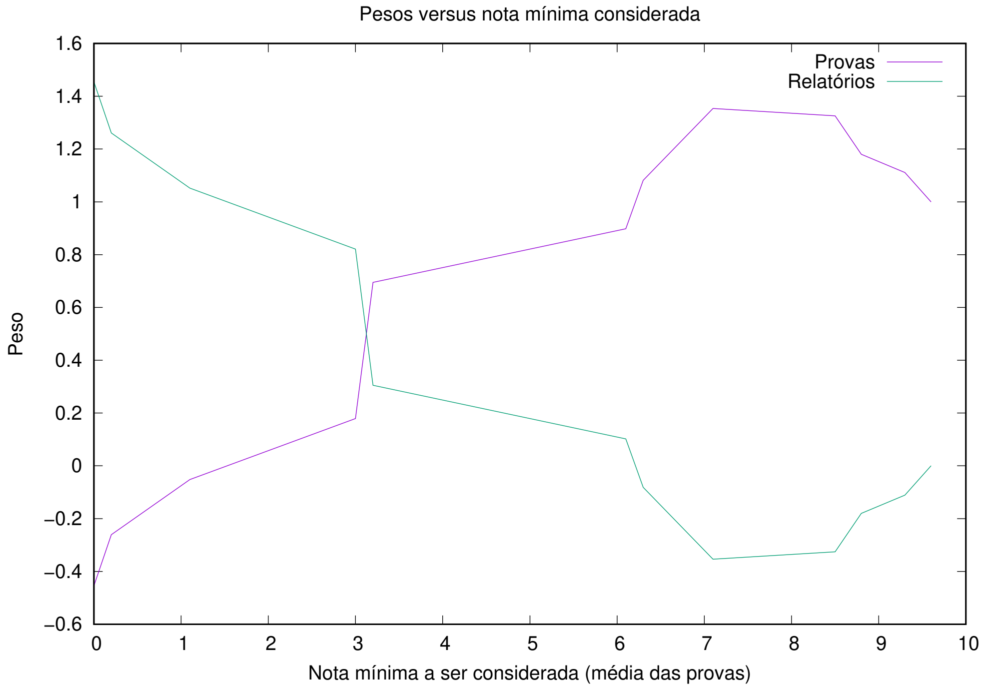

Figura 2: Pesos para a média das provas (em roxo) e dos relatórios (em verde), em função da nota mínima considerada para o método de mínimos quadrados sem a barreira de domínio. Os valores numéricos dos pontos presentes nesse gráfico podem ser vistos na Tabela 1.

	Pesos			
Nota de corte	Prova 1	Prova 2	Relatório 1	Relatório 2
0.0	0.138600	-0.4147890	0.8224170	0.4537720
0.2	1.141850	-0.7449710	-0.3473890	0.9505130
1.1	1.666390	-0.7560880	-0.7351380	0.8248360
3.0	1.311960	-0.4370530	-0.8058130	0.9309100
3.2	1.101590	-0.1359640	-0.7549420	0.7893160
6.1	1.548200	-0.3045360	-0.8263390	0.5826740
6.3	1.321180	-0.0287221	-0.7614710	0.4690170
7.1	0.672257	0.6612190	-0.3681950	0.0347192
8.5	0.420290	0.8608700	-0.1673910	-0.1137680
8.8	0.678161	0.4482760	0.0775862	-0.2040230
9.3	0.000000	1.0000000	0.0000000	0.0000000

Tabela 2: Tabela de pesos obtidos pela configuração que leva em conta cada avaliação independentemente. Note que, diferentemente da Tabela 1, em nenhuma das notas de corte há apenas pesos positivos. Isso ilustra a dificuldade de aplicação desse método para a maximização no caso em que há múltiplos pesos independentes. Um gráfico dos dados dessa tabela pode ser visto na Figura 3.

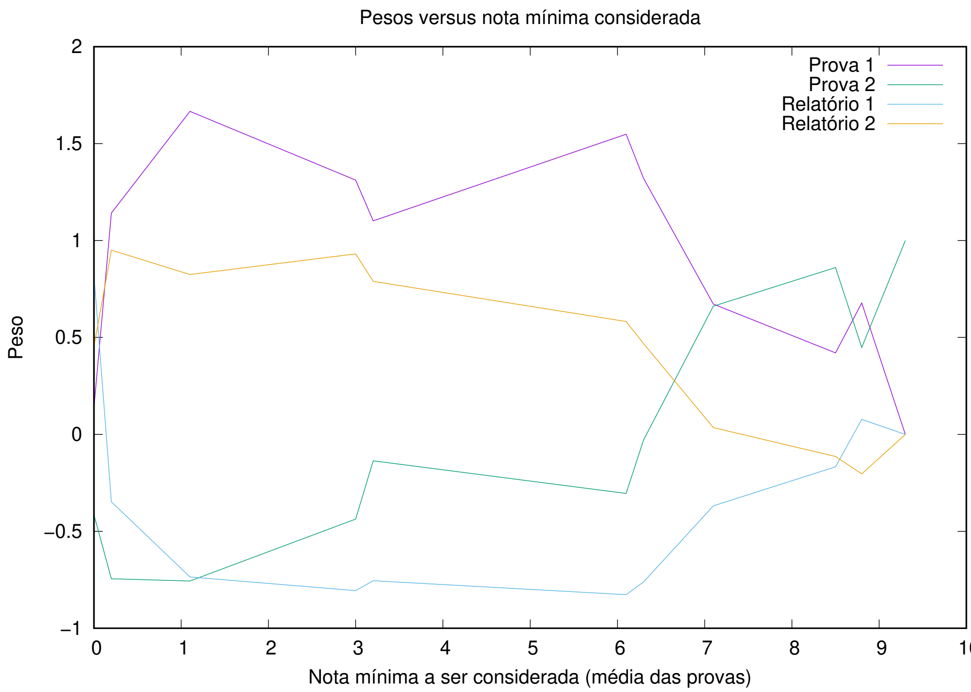

Figura 3: Pesos para a prova 1 (em roxo), para a prova 2 (em verde), para o relatório 1 (em azul) e para o relatório 2 (em amarelo) em função da nota mínima considerada para a média aritmética das provas. Os dados numéricos desse gráfico podem ser vistos na Tabela 2.

sugere que, com mais pesos, a aplicação do método de mínimos quadrados exigindo restrições de domínio pode ser cada vez mais difícil de fornecer resultados frutíferos.

Seguiu-se então uma análise do estado final da turma em termos de estudantes aprovados, em exame final ou reprovados. Os resultados numéricos dessa análise estão, para ambas as configurações e levando em conta tanto os casos com e sem barreira de domínio, na Tabela 3 e nos gráficos das Figuras 4 (pesos simples com barreira de domínio), 5 (pesos simples sem barreira de domínio) e 6 (pesos independentes).

Nos gráficos das Figuras 4, 5 e 6, e também na Tabela 3, é possível observar que a configuração que atinge o maior número de aprovados é a média com barreira de domínios. Entretando, as duas outras são capazes (com pesos negativos) de encontrar configurações tais que nenhum aluno estaria imediatamente reprovado. É interessante observar também que, quando a nota de corte se torna maior do que zero, e os alunos com notas nulas nas provas passam a ser desconsiderados, há, na Figura 6, um aumento no número de aprovações, o que mostra que esses alunos são têm notas que destoam do padrão da turma. Foram destacadas as linhas que também estão destacadas na Tabela 1, nas quais os pesos da configuração de pesos simples se apresentaram positivos.

Outro fator a ser observado nesses gráficos é de que a maior liberdade para os pesos não implica necessariamente numa maior capacidade de aprovação, isto é, há mais notas altas na configuração de pesos simples com barreira de domínio.

Finalmente, estudou-se também como se comporta a média da turma ao variarmos o limiar de nota. Os resultados estão na Tabela 4 e no gráfico da Figura 7.

Observa-se na Tabela 4 e na Figura 7 que há uma vantagem com o uso dos pesos independentes: a média da turma se mantém significativamente acima quando comparada aos outros dois casos. De forma similar, a média da turma é superior quando os pesos não passam por uma barreira de domínio do que quando passam. Esse padrão continua até a região final do gráfico, quando são consideradas apenas notas acima de 6, e a ordem dos gráficos passa a se inverter.

	Médio com BD			Médio sem BD			Independentes		
Nota de corte	Apr.	Final	Rep.	Apr.	Final	Rep.	Apr.	Final	Rep.
0.0	13	1	1	10	5	0	11	4	0
0.2	13	1	1	10	5	0	12	2	1
1.1	13	1	1	9	5	1	11	3	1
3.0	11	3	1	11	3	1	11	2	2
3.2	9	4	2	9	4	2	11	2	2
6.1	9	2	4	9	2	4	10	3	2
6.3	9	2	4	7	4	4	10	3	2
7.1	9	2	4	7	4	4	9	2	4
8.5	9	2	4	7	4	4	7	4	4
8.8	9	2	4	7	4	4	7	4	4
9.3	9	2	4	7	4	4	7	4	4
9.6	9	2	4	9	2	4			

Tabela 3: Tabela de estado final dos alunos da disciplina em termos de alunos aprovados, em exame final ou reprovados para todas as configurações de nota analisadas. Observe que a última linha da coluna na configuração de pesos independentes está faltando. Isso ocorre pois, por possuir três valores desconhecidos, o método de mínimos quadrados não pode ser aplicado quando a matriz das diferenças das notas (ver Equação 5) possui menos de três linhas. À esta tabela estão associados os gráficos das Figuras 4, 5 e 6. As linhas destacadas correspondem às linhas também destacadas na Tabela 1, onde todos os pesos da configuração de pesos simples se apresentaram positivos.

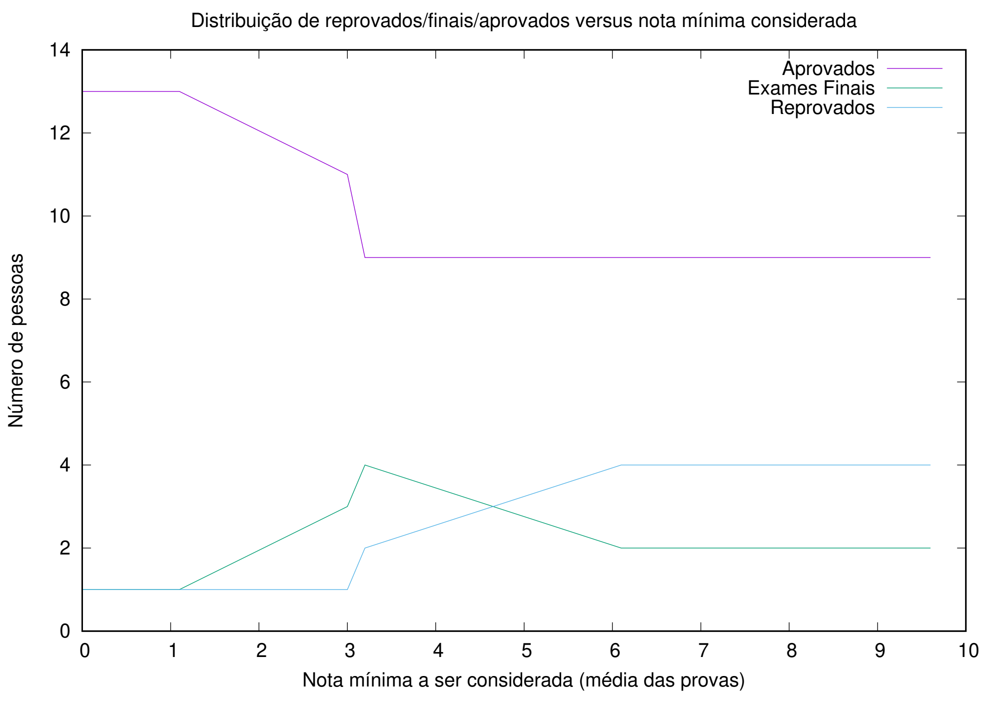

Figura 4: Gráfico do número de alunos aprovados (em roxo), em exame final (em verde) e reprovados (em azul) para a configuração de pesos simples, que leva em conta a nota média dos relatórios e a nota média da prova, com barreira de domínio, em função da nota de corte mínima considerada para o método de mínimos quadrados.

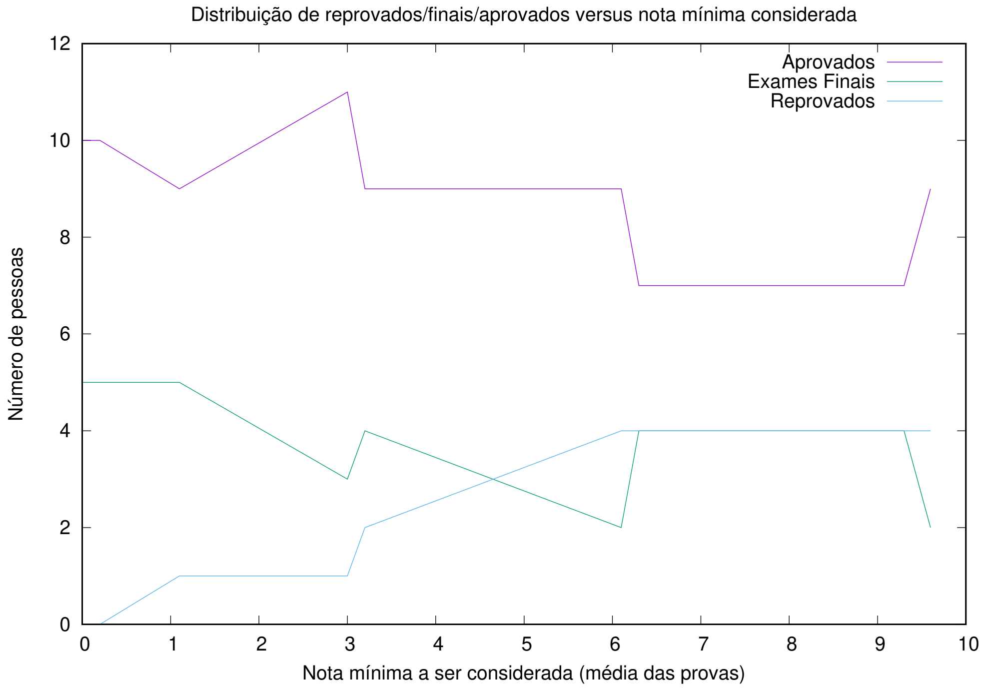

Figura 5: Gráfico do número de alunos aprovados (em roxo), em exame final (em verde) e reprovados (em azul) para a configuração de pesos simples, que leva em conta a nota média dos relatórios e a nota média da prova, sem barreira de domínio, em função da nota de corte mínima considerada para o método de mínimos quadrados.

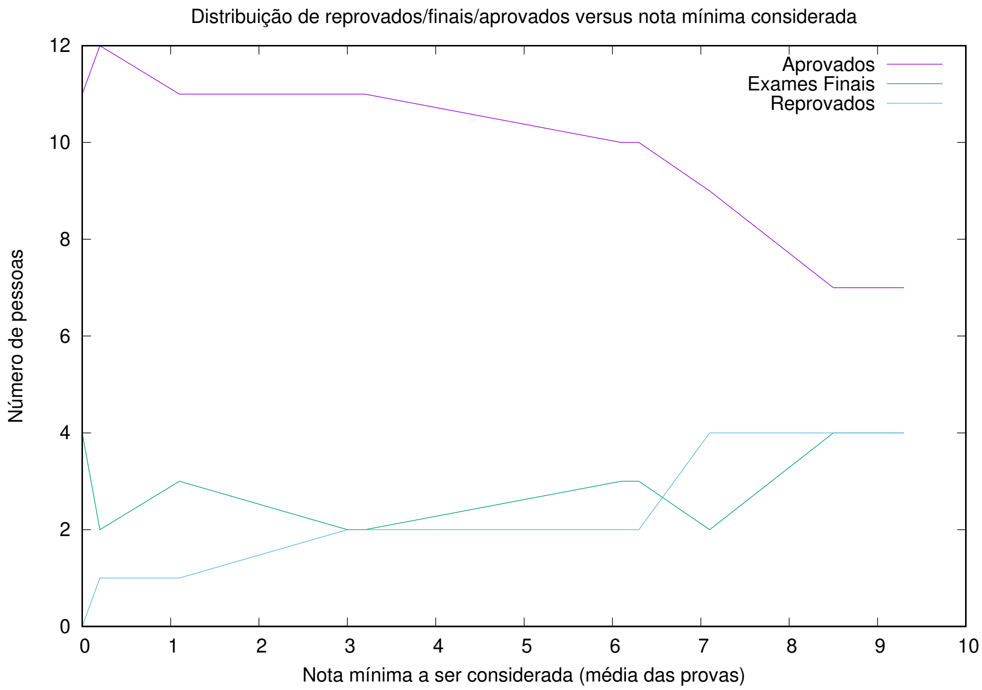

Figura 6: Gráfico do número de alunos aprovados (em roxo), em exame final (em verde) e reprovados (em azul) para a configuração de pesos independentes, em função da nota de corte mínima considerada para o método de mínimos quadrados.

	Método		
Nota de corte	Média com BD	Média sem BD	Independentes
0.0	7.20000	7.50143	7.54987
0.2	7.20000	7.37301	7.84565
1.1	7.20000	7.23458	7.74893
3.0	7.08109	7.08109	7.53332
3.2	6.73897	6.73897	7.24801
6.1	6.60445	6.60445	7.28547
6.3	6.53667	6.48256	7.03320
7.1	6.53667	6.30198	6.37349
8.5	6.53667	6.32048	6.18549
8.8	6.53667	6.41722	6.50490
9.3	6.53667	6.46296	6.14667
9.6	6.53667	6.53667	

Tabela 4: Média da turma em função do limiar de nota. As linhas destacadas correspondem às linhas destacadas na Tabela 1, quando os pesos da configuração de pesos simples se mostraram positivos. O gráfico correspondente a essa tabela pode ser visto na Figura 7.

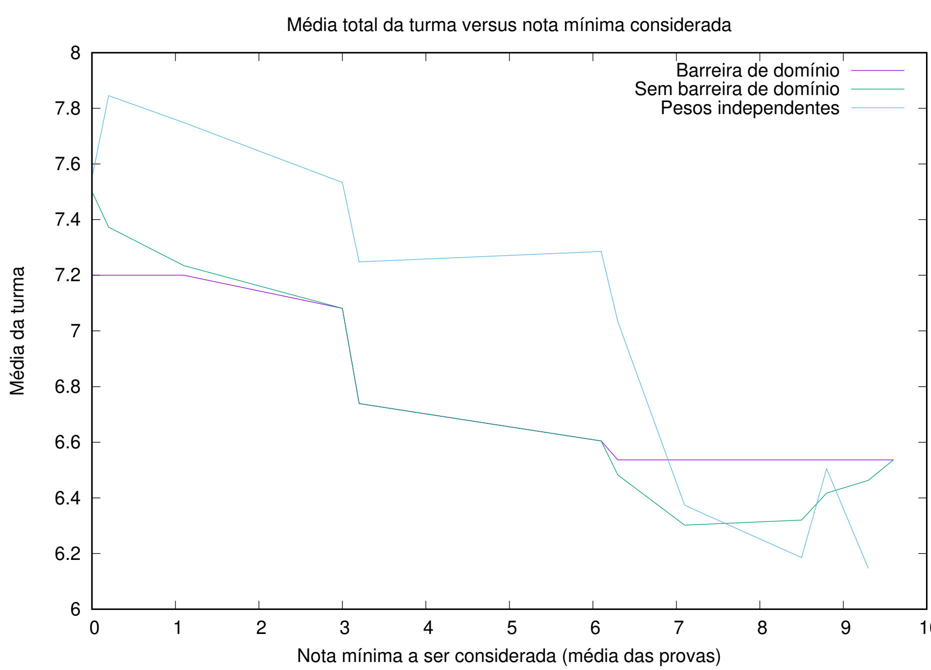

Figura 7: Gráfico da média da turma para a configuração de pesos simples com barreira de domínio (em roxo), sem barreira de domínio (em verde) e com pesos independentes (em azul) em função da nota mínima a ser considerada. Os valores numéricos dos pontos nesse gráfico estão na Tabela 4.

Isso pode ser atribuído à redução do número de elementos para os quais a maximização está sendo feita, o que transmite aos pesos detalhes específicos de um único aluno, ao invés de considerar um padrão mais geral para a turma. Em especial, é possível notar pelo gráfico de pesos independentes uma "volatilidade" maior, especialmente pela sua queda brusca quando a nota mínima passa de 6.0 para 7.0 e pelo pico entre as notas de 8.0 e 9.0, que; em consonância com a hipótese de que a redução do número de alunos para o qual a maximização é feita transmite ao resultado final detalhes individuais que não necessariamente correspondem a um padrão real no grupo todo; parece se dever a um aluno cuja configuração de notas destoa da turma, e que possui uma média das provas nessa faixa, e, quando removido, causa um salto na média da turma.

5 Conclusões

Por meio dos resultados obtidos nessa aplicação empírica do método dos mínimos quadrados, constatou-se sua validade para situações desse contexto e de contexto semelhante.

Valem, todavia, as ressalvas já apontadas no corpo do texto, em especial da dificuldade de estabelecer pesos válidos para o método quando há muitas variáveis independentes. Uma solução possível porém não empregada para esse problema seria uma extensão da barreira de domínio para que, quando um dos pesos fosse negativo ou nulo, desconsiderasse-se sua nota correspondente e fosse refeito com as demais notas o método de maximização levando em conta uma matriz de dimensões reduzidas.

Contudo, o método de mostrou frutífero quando consideramos os pesos apenas para as médias aritméticas das provas e dos relatórios, para os quais, de acordo com a Tabela 1, há três estados (linhas destacadas) com ambos os pesos válidos. Algum desses pesos poderá então ser usado realmente na situação real para o cálculo das médias finais dos alunos.

Curiosamente, na disciplina de Física Básica 2, a porcentagem do peso da média das provas e a porcentagem do peso da média dos relatórios usualmente utilizados é 70% e 30%, respectivamente, o que está muito próximo do valor central entre os três valores destacados na Tabela 1. Isso sugere um possível padrão que poderá, em algum estudo futuro, ser analisado.

Por fim, ressalta-se que o trabalho atingiu seu objetivo de atuar como *proof of concept* para a ideia de aplicação do método dos mínimos quadrados para maximização, na média, das notas de alunos de uma disciplina.