

Regularization, Shrinkage & Cellular Bioenergetics

Amostra combinada: intuição em português, vocabulário técnico e fórmulas prontas para impressão A4

Fontes: Introduction to Statistical Learning (Cap. 6) & Campbell Biology (Cap. 9) · Modo Bilíngue (PT + EN)

COMO ESTUDAR COM ESTE GUIA

1. Leia uma vez do início ao fim para construir a intuição, sem tentar decorar fórmulas.
2. Feche o PDF e responda ao quiz de recuperação ativa na plataforma.
3. Revise os flashcards priorizados pelos seus erros ou exporte o baralho para o Anki.

PARTE 1 – REGULARIZAÇÃO E ENCOLHIMENTO (MODO BILÍNGUE)

1 Ridge regression (L2) MODO BILÍNGUE

A INTUIÇÃO PRÁTICA

Ridge é um imposto progressivo: todo coeficiente paga alguma coisa, os maiores pagam mais, e ninguém é demitido. No fim todo mundo continua na folha de pagamento, só que ganhando menos.

- **Penalty** — adds the squared L2 norm of the coefficients to the residual sum of squares.
- **Effect** — shrinks coefficients toward zero continuously, but never sets them exactly to zero.
- **Tuning** — controlled by a single hyperparameter $\lambda \geq 0$, selected via cross-validation.

$$\hat{\beta}^{\text{ridge}} = \arg \min_{\beta} \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p \beta_j^2$$

PRA FIXAR

O mesmo mecanismo aparece com dois nomes na literatura: em estatística chama-se **ridge**, em deep learning chama-se **weight decay**. A atualização do gradiente multiplica o peso por $(1 - \eta\lambda)$ a cada passo — daí o nome "decay": o peso encolhe uma fração a cada iteração. Perceber que são a mesma ideia economiza metade do esforço de estudo.

Conexão entre capítulos: o parâmetro de complexidade α na poda de árvores cumpre exatamente o mesmo papel que λ na regularização — um preço cobrado por cada grau extra de complexidade.

2

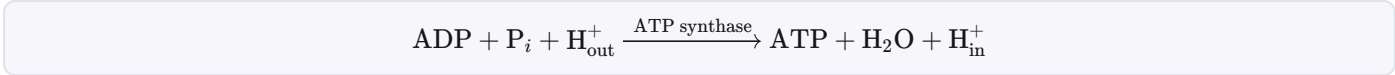
Cellular respiration & ATP synthesis

CAPÍTULO 1 • CÉLULAS

A INTUIÇÃO PRÁTICA

A glicose é como uma nota de alto valor e o ATP é o trocado no bolso. A célula não consegue pagar reações químicas direto com uma molécula inteira de glicose; a mitocôndria funciona como a casa de câmbio que converte essa reserva em dezenas de moedas prontas para uso imediato.

- **Proton-motive force** — electrochemical gradient of H^+ pumped across the inner mitochondrial membrane.
- **Chemiosmosis** — protons flow back through ATP synthase, driving the phosphorylation of ADP into ATP.



3

Revisão de prova: Which penalty can set a coefficient to exactly zero?

ANSWER

✓ **Lasso (L1)**. Ridge shrinks coefficients asymptotically toward zero, whereas Lasso performs variable selection.

THEORY

The L1 constraint region $\sum |\beta_j| \leq s$ is a diamond with sharp corners on the coordinate axes, so the elliptical contours of the RSS loss typically intersect a corner first — where one or more coordinates equal zero. The L2 region $\sum \beta_j^2 \leq s$ is a smooth sphere with no corners, so tangency almost never lands on an axis.

PEGADINHA CLÁSSICA

"Regularização L2 zera variáveis irrelevantes" é falso. L2 apenas reduz a magnitude; apenas L1 produz modelos esparsos.

Método	Penalidade adicionada	Quando escolher na prática
OLS (Mínimos Quadrados)	Nenhuma ($\lambda = 0$)	Poucos preditores, baixa colinearidade, $n \gg p$
Ridge (L2)	$\lambda \sum \beta_j^2$ (encolhimento proporcional)	Muitos preditores correlacionados que contribuem juntos
Lasso (L1)	$\lambda \sum \beta_j $ (seleção esparsa)	Apenas um subconjunto pequeno de variáveis importa
Elastic Net	Combinação L1 + L2	Grupos de variáveis correlacionadas que devem entrar ou sair juntos