Clase 075 — Regresión logística binaria y softmax

Parte: 1 — Machine Learning Clásico · Fuente: Géron, cap. 4. Duración estimada: 70 min.

Clase 075 — Regresión logística binaria y softmax

> *Parte: 1 — Machine Learning Clásico · Fuente: Géron, cap. 4. Duración estimada: 70 min.*

Objetivo

Que el alumno entienda la regresión logística como modelo lineal para clasificación: cómo la sigmoide convierte un score lineal en probabilidad, por qué se entrena minimizando log-loss (cross-entropy), y cómo se generaliza a multiclase con softmax. Además, que sepa diagnosticar si las probabilidades que devuelve un clasificador están bien calibradas y cómo corregirlas si no.

Resultados de aprendizaje

Al finalizar la clase, el alumno podrá:

1. Derivar la sigmoide σ(z) = 1/(1+e^-z) como puente entre score lineal y probabilidad, y explicar por qué no se usa MSE en clasificación.
2. Entrenar LogisticRegression binaria de sklearn, interpretar coeficientes como log-odds y la frontera de decisión.
3. Extender a multiclase con multi_class='multinomial' (softmax) y diferenciar de 'ovr' (one-vs-rest).
4. Evaluar con log-loss y Brier score, no solo accuracy.
5. Diagnosticar y corregir calibración con calibration_curve y CalibratedClassifierCV (Platt / isotonic).

Temas

#	Tema	Por qué importa
1	Sigmoide y log-odds	Conecta regresión lineal con probabilidad
2	Log-loss (cross-entropy binaria)	Función de costo convexa, derivable, penal
3	Regularización (C, penalty)	sklearn regulariza por default — C=1/λ.
4	Softmax para multiclase	Generaliza sigmoide a K clases con probabi
5	multinomial vs ovr	El primero es softmax real; el segundo ent
6	Predict_proba y calibración	El score no siempre es probabilidad confia

Versión profundizada — 2026

El tema moderno que antes vivía como complemento dentro de esta clase ahora tiene su(s) clase(s) propia(s) con patrón completo, ejercicios y homework:

- Clase 067a — Calibración de probabilidades: Platt, isotonic, temperature scaling

Definiciones y características

Sigmoide σ(z)

: Función 1/(1+exp(-z)) que mapea → (0,1). Su inverso es el logit log(p/(1-p)). La regresión logística modela logit(p) = w·x + b (linealidad en los log-odds).

Log-loss (cross-entropy binaria)

: -[y·log(p) + (1-y)·log(1-p)]. Función de costo convexa de la logística. Penaliza fuerte la confianza alta cuando te equivocás (predecir 0.99 y que sea 0 cuesta ~4.6).

Softmax

: Generalización de la sigmoide a K clases: softmax(z_k) = exp(z_k) / Σ exp(z_j). Probabilidades positivas que suman 1. Es la activación final de logística multinomial y de la última capa en clasificadores neuronales.

Cross-entropy categórica

: Generalización del log-loss a K clases: -Σ_k y_k · log(p_k) con y one-hot. Pareja natural de softmax.

Calibración

: Propiedad de que P(y=1 | ŷ=p) ≈ p para todo p. Un modelo calibrado al 0.7 acierta como positivo el 70% de las veces que dice "0.7".

Reliability diagram

: Gráfico de fracción observada de positivos vs score promedio en bins. Diagonal = perfecto. Curva en S = típica de árboles; curva inversa = sobre-confianza.

Brier score

: mean((p - y)^2). MSE entre probabilidades y labels 0/1. Resume calibración + discriminación en un escalar. Menor = mejor.

CalibratedClassifierCV

: Wrapper de sklearn que envuelve un clasificador base y calibra sus probabilidades vía cross-validation interno. Métodos: 'sigmoid' (Platt) o 'isotonic'.

Dataset / recursos

- Iris (3 clases) para softmax — sklearn.datasets.load_iris().
- Breast cancer Wisconsin (binario) para logística y calibración — load_breast_cancer().
- Sintético desbalanceado con make_classification(weights=[0.9, 0.1]) para visualizar mis-calibración de un RandomForest.

Ejercicios

1. Logística binaria desde cero. Entrená LogisticRegression() sobre breast cancer. Reportá accuracy, log-loss y matriz de confusión. Imprimí los 5 coeficientes con mayor |w| e interpretá uno como odds-ratio (exp(w)).

2. Frontera de decisión. Con 2 features de iris (solo 2 clases primero), graficá la frontera lineal de la logística y los puntos. Cambiá C entre 0.01 y 100 y observá cómo la frontera se vuelve más/menos rígida.

3. Softmax sobre iris. LogisticRegression(multi_class='multinomial', solver='lbfgs'). Comparalo con multi_class='ovr' en log_loss y accuracy. Imprimí predict_proba de 3 muestras y verificá que sumen 1.

4. Reliability diagram de un RandomForest. Entrená RandomForestClassifier(n_estimators=100) sobre el dataset sintético desbalanceado. Computá calibration_curve con n_bins=10 y graficá vs diagonal. Reportá Brier score.

5. Calibrar con CalibratedClassifierCV. Sobre el mismo RF: aplicá method='sigmoid' y method='isotonic' (cv=5). Re-graficá los tres reliability diagrams (RF crudo, +Platt, +isotonic) y compará Brier scores. Reportá cuál calibra mejor y por qué te parece.

Homework verificable

Notebook que sobre make_classification(n_samples=20000, weights=[0.9, 0.1], random_state=42): (a) entrena LogisticRegression y RandomForestClassifier; (b) reporta accuracy, log-loss y Brier score de ambos en test; (c) grafica reliability diagram de ambos en la misma figura; (d) calibra el RF con CalibratedClassifierCV(method='isotonic', cv=5) y reporta el nuevo Brier; (e) escribe 2-3 líneas interpretando: ¿quedó el RF mejor calibrado que la logística cruda?

Criterio de aceptación: Brier score del RF calibrado debe ser menor que el del RF crudo. El reliability diagram del calibrado debe estar visiblemente más cerca de la diagonal.

Errores comunes

Síntoma / mensaje	Causa y cómo arreglar
Usar predict_proba directamente para tomar	Asumir que el score es probabilidad. Fix:
ConvergenceWarning: lbfgs failed to conver	Features sin escalar o max_iter bajo. Fix:
Coeficientes enormes con C muy alto en dat	Sin regularización, la logística diverge c
multi_class='ovr' y log-loss raro en multi	OvR entrena K binarios independientes — la
Calibrar sobre el mismo set de entrenamien	Leakage — los scores ya están sobreajustad

Preguntas frecuentes

¿Por qué log-loss y no MSE para clasificación?

Con MSE sobre una sigmoide la superficie de costo es no convexa (varios mínimos locales) y los gradientes se saturan en los extremos. Log-loss + sigmoide da costo convexo y gradientes limpios (p - y).

¿Platt o isotonic?

Platt si tenés ~1000 ejemplos de calibración y la curva de mis-calibración parece sigmoidea (típico SVM, NN pequeñas). Isotonic si tenés ~10k+ y/o la mis-calibración no es monótona-sigmoidea (RF, boosting). Regla práctica: probá ambos en validación y quedate con el de menor Brier.

¿predict_proba de LogisticRegression está calibrado?

Generalmente sí, porque la logística optimiza log-loss directamente — sus probabilidades suelen ser razonables. Igual chequealo con calibration_curve antes de usarlas para decisiones críticas; regularización fuerte (C chico) puede sub-confiar el output.

¿Softmax y sigmoide son lo mismo en binario?

Equivalentes: softmax con K=2 colapsa a sigmoide. sklearn con multi_class='multinomial' y 2 clases hace lo mismo que el modo binario por default.

¿Calibrar afecta el accuracy o solo las probabilidades?

Platt e isotonic son monótonas no-decrecientes → no cambian el ranking ni el argmax → accuracy y AUC quedan iguales. Lo que cambia es el log-loss, el Brier score, y el threshold óptimo cuando elegís uno distinto de 0.5.

Referencias

- Géron, cap. 4 § Logistic Regression y § Softmax Regression.
- sklearn — Probability calibration
- sklearn — LogisticRegression
- Niculescu-Mizil & Caruana (2005), Predicting Good Probabilities With Supervised Learning, ICML — paper canónico sobre Platt vs isotonic en RF, SVM, boosting, NB.
- Guo et al. (2017), On Calibration of Modern Neural Networks — temperature scaling.

Siguiente clase

Clase 076 — Calibración de probabilidades: Platt, isotonic, temperature scaling

Apéndice: notebook (primer bloque)

Primera celda ejecutable del notebook de la clase.

```
# Imports y configuración inicial
```

Archivos complementarios

- notebook.ipynb