# Métricas de Avaliação em Classificação com CNNs

Este documento explica as principais métricas usadas para avaliar o desempenho de modelos de classificação, como redes neurais convolucionais (CNNs), usando como exemplo o conjunto de dados CIFAR-10.

#### **♦ Matriz de Confusão**

A matriz de confusão mostra os acertos e erros do modelo comparando a classe verdadeira com a classe prevista.

- Cada linha representa a classe \*\*verdadeira\*\*.
- Cada coluna representa a classe \*\*prevista\*\*.
- A diagonal principal mostra os acertos.
- Os valores fora da diagonal são erros de classificação.

Por exemplo, se uma imagem de 'gato' for classificada como 'cachorro', isso será registrado fora da diagonal.

#### **Q** Como interpretar a matriz de confusão

A matriz permite visualizar onde o modelo está errando sistematicamente. Se, por exemplo, a linha 'cat' tem muitos valores nas colunas 'dog' e 'bird', isso indica que o modelo confunde essas classes com frequência.

Isso pode significar que as classes têm características visuais semelhantes, ou que há poucos exemplos no dataset para que o modelo aprenda a diferenciá-las.

Use a matriz para saber quais classes precisam de mais atenção, mais dados, ou ajustes no modelo.

## **Métricas por Classe**

Essas métricas são calculadas individualmente para cada classe:

- Precisão (Precision)
- Mede quantas das amostras que o modelo disse ser de uma classe realmente são dessa classe.
- Alta precisão significa poucos \*\*falsos positivos\*\*.

Exemplo: O modelo disse que 100 imagens são de 'gato', mas só 80 eram mesmo. A precisão é 80%.

- **V** Recall
- Mede quantas amostras de uma classe foram corretamente identificadas pelo modelo.
- Alto recall significa poucos \*\*falsos negativos\*\*.

Exemplo: De 100 imagens reais de 'gato', o modelo encontrou 70. O recall é 70%.

- F1-score
- Média harmônica entre precisão e recall.
- Equilibra os dois valores para dar uma visão geral.
- Ideal quando há desequilíbrio entre classes ou quando falsos positivos e negativos têm impacto.

Fórmula: F1 = 2 \* (precisão \* recall) / (precisão + recall)

- Support
- Indica quantos exemplos reais da classe existem no conjunto de teste.
- Serve como base para entender o peso de cada classe no resultado final. Exemplo: A classe 'avião' tem support  $1000 \rightarrow$  significa que há 1000 imagens de avião no teste.

## Interpretação do F1-score

Use a tabela abaixo para interpretar o desempenho do modelo com base no F1-score:

Intervalo do F1-score	Desempenho
0.90 – 1.00	Excelente
0.75 – 0.89	Bom
0.50 - 0.74	Razoável / Médio
< 0.50	Fraco – Modelo precisa de ajustes

#### **Quando o F1-score é mais importante?**

- Quando você não quer apenas acertar muito, mas acertar de forma confiável.
- Quando o conjunto de dados é desbalanceado (ex: muito mais 'gato' do que 'avião').
- Em casos onde falsos positivos e falsos negativos têm custo alto, como:
- Detecção de doença
- Segurança
- Fraude bancária

## ☆ Conclusão

Essas métricas ajudam a identificar se o modelo está aprendendo corretamente, se está generalizando bem e quais classes estão sendo confundidas. Use a matriz de confusão para encontrar padrões de erro, e combine precisão, recall e F1-score para decisões mais informadas sobre ajustes no modelo.

## Interpretação Visual de uma Matriz de Confusão

Abaixo temos uma estrutura simplificada de uma matriz de confusão para facilitar o entendimento dos acertos e erros do modelo.

#### Estrutura típica de uma matriz (simplificada)

	Pred: Cat	Pred: Dog	Pred: Frog
Real: Cat	80	10	10
Real: Dog	15	70	15
Real: Frog	5	10	85

#### Como ler:

- A diagonal (80, 70, 85) representa os acertos.
- Os outros valores representam os erros de classificação.
- Ex: O modelo errou 10 imagens de gato e chamou de dog.
- Ex: Ele errou 5 imagens de sapo (frog) e chamou de gato.

### **©** Como interpretar no CIFAR-10

No CIFAR-10, temos 10 classes:

['airplane', 'automobile', 'bird', 'cat', 'deer',

'dog', 'frog', 'horse', 'ship', 'truck']

Portanto, a matriz de confusão real terá dimensão 10x10.

- Cada linha representa as imagens reais (rótulos do dataset)
- Cada coluna representa o que o modelo previu
- A diagonal principal mostra os acertos
- Os valores fora da diagonal são os erros

### **Exemplo de análise real:**

Se a linha "cat" da matriz estiver assim:

Real: Cat  $\rightarrow$  Pred: Cat = 600, Dog = 250, Bird = 100, Outras = 50

Significa que o modelo acertou 600 gatos, mas errou 400:

- 250 ele achou que eram cachorros
- 100 achou que eram pássaros
- 50 foram confundidos com outras classes

#### **✓** Como usar a matriz:

- Identificar classes que o modelo erra muito
- Ver se há confusões sistemáticas (ex: sempre troca truck por automobile)
- Guiar ajustes no modelo, como:
  - Aumentar dados de uma classe específica
  - - Usar data augmentation em classes que erram mais

• - Fazer fine-tuning com pesos balanceados

## **Estrutura típica de uma matriz (simplificada)**

Imagine que temos 3 classes: cat, dog, frog. A matriz de confusão pode ser representada assim:

	Pred: Cat	Pred: Dog	Pred: Frog
Real: Cat	80	10	10
Real: Dog	15	70	15
Real: Frog	5	10	85

## Como ler:

- • A diagonal (80, 70, 85) → são os \*\*acertos\*\*
- • Os outros valores → são os \*\*erros\*\*
  - - Ex: O modelo \*\*errou 10 imagens de gato\*\* e chamou de \*\*dog\*\*
  - - Ex: Ele \*\*errou 5 imagens de sapo (frog)\*\* e chamou de \*\*gato\*\*