

Makine Ogrenmesi Model Degerlendirme Metrikleri Ozeti

Confusion Matrix (Karmasiklik Matrisi)

Ikili siniflandirmada (ornegin: spam / spam degil) tahminler su sekilde kategorize edilir:

- TP (True Positive): Pozitifi dogru tahmin ettin.
- FP (False Positive): Negatifi yanlis pozitif tahmin ettin.
- FN (False Negative): Pozitifi yanlis negatif tahmin ettin.
- TN (True Negative): Negatifi dogru tahmin ettin.

Bu degerler uzereinden diger metrikler hesaplanir. Ornegin accuracy, precision, recall ve F1-score.

Classification Metrikleri (Siniflandirma)

Bu metrikler modelin ne kadar basarili oldugunu gosterir:

- Accuracy: Genel dogruluk orani. Dengeli veri setlerinde kullanisli.
- Precision: Pozitif tahminlerin ne kadar dogru oldugunu gosterir.
- Recall (Sensitivity): Gercek pozitiflerin ne kadarini yakalayabildigini gosterir.
- F1-Score: Precision ve Recall arasindaki dengeyi olcer.
- Specificity: Gercek negatiflerin dogru bulunma orani.
- ROC-AUC: Modelin ayiricilik gucu.
- Log Loss: Olasilik tahminlerinin dogrulugunu olcer.
- MCC: Dengesiz veri setlerinde guvenilir bir olcum.

Regression Metrikleri

- MSE: Ortalama kare hata. Kucukse iyi.
- RMSE: Hatalarin karekok ortalamasi.
- MAE: Ortalama mutlak hata.
- MAPE: Yuzdesel hata.
- R^2 : Aciklanan varyans orani, 1'e ne kadar yakin o kadar iyi.

Clustering Metrikleri (Kumeleme)

- Silhouette Score: Kume icindeki benzerlik ve kume disindaki farklilik. +1'e yakin olmasi iyi.
- Davies-Bouldin Index: Kucuk olmasi daha iyi.

- Calinski-Harabasz Index: Büyük degerler daha iyi kumeleme gosterir.

Recommendation / Ranking Metrikleri

- Precision@K: İlk K öneride doğru olanların oranı.

- Recall@K: Tüm doğru item'lerin ne kadarını yakaladı.

- MAP: Ortalama kesinlik.

- NDCG: Doğru itemlerin ne kadar üst sıralarda çıktığını gösterir.

Classification

| Metric | Ideal Value | Formula | Comment |
|-----------------------------|-------------|---|---|
| Accuracy | 1 | $(TP+TN) / (TP+TN+FP+FN)$ | Overall correctness. Reliable on balanced datasets. |
| Precision | 1 | $TP / (TP+FP)$ | How correct are positive predictions? |
| Recall (Sensitivity) | 1 | $TP / (TP+FN)$ | How many true positives are captured? |
| F1-Score | 1 | $2 * (Precision*Recall)/(Precision+Recall)$ | Balance between Precision and Recall. |
| Specificity | 1 | $TN / (TN+FP)$ | Correctly identifying true negatives. |
| ROC-AUC | 1 | - | 0.5 = random, 1 = perfect. |
| Log Loss | 0 | $-1/N \sum [y \log(p) + (1-y) \log(1-p)]$ | Lower is better for probability estimates. |
| MCC | +1 | $(TP*TN - FP*FN)/\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}$ | -1 = worst, 0 = random, +1 = perfect. |

Regression

| Metric | Ideal Value | Formula | Comment |
|------------|-------------|----------------------------|---------------------|
| MSE | 0 | $1/N \sum (y - \hat{y})^2$ | Mean squared error. |

| | | | |
|----------------------|----|---|--|
| | | | Lower is better. |
| RMSE | 0 | $\sqrt{\text{MSE}}$ | Root mean squared error. |
| MAE | 0 | $1/N \sum y - \hat{y} $ | Mean absolute error. |
| MAPE | 0% | $100/N \sum (y - \hat{y})/y $ | Mean absolute percentage error. Lower is better. |
| R² | 1 | $1 - \frac{\sum (y - \hat{y})^2}{\sum (y - \bar{y})^2}$ | Explained variance ratio. Closer to 1 is better. |

Clustering

| Metric | Ideal Value | Formula | Comment |
|--------------------------|-------------|---|---|
| Silhouette Score | 1 | $(b-a)/\max(a,b)$ | 1 = well separated, 0 = overlapping, -1 = wrong clustering. |
| Davies-Bouldin | 0 | - | Lower is better. |
| Calinski-Harabasz | High | $\text{Tr}(B_k)/\text{Tr}(W_k) * (N-k)/(k-1)$ | Higher = better clustering. |

Recommendation / Ranking

| Metric | Ideal Value | Formula | Comment |
|--------------------|-------------|--|---|
| Precision@K | 1 | Correct recommendations / K | Accuracy of the top-K recommendations. |
| Recall@K | 1 | Correct recommendations / All relevant items | How many relevant items are retrieved? |
| MAP | 1 | $1/N \sum \text{Average Precision}$ | Mean average precision. |
| NDCG | 1 | DCG / IDCG | Higher if relevant items are ranked higher. |