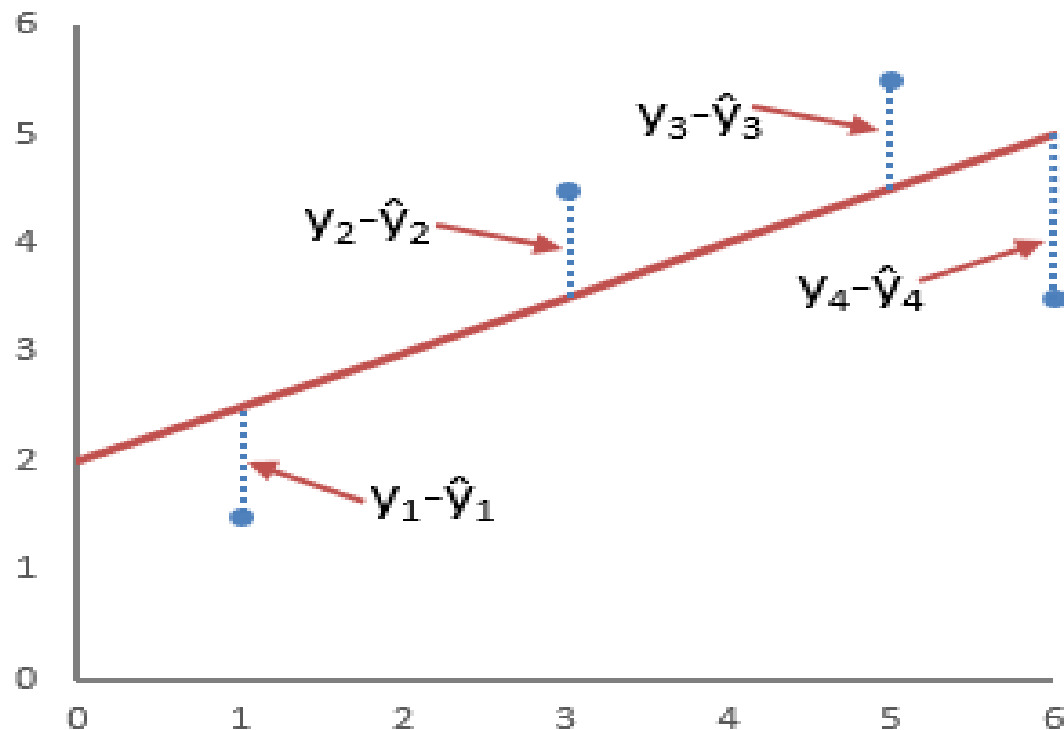


Машинное обучение

Лекция 4. Функции потерь и метрики качества

Пример. Приближение с помощью кривых



Функция потерь

$$\hat{y} = f(x, w)$$

$$L(w) \rightarrow \min$$

Пример. Оценка параметра

- Метод максимального правдоподобия
- Оценка апостериорного максимума
- Байесовская оценка

Примеры

- Построение функции Ляпунова
- Оптимальное управление

Ещё пример

Вахтер хочет понять кого пускать в парадную. Он хочет минимизировать свою работу (больше спать) по:

- проверке входящих;
- разборкам с жильцами/руководством;
- уборке/проветриванию.

Для этого ему надо проверять входящих (думать). Однако, минимизировать “время сна” напрямую очень сложно. Задача помочь бедному вахтеру.

Желаемые свойства

- Выпуклость: легко оптимизировать.
- Дифференцируемость: производная по параметрам модели существует и непрерывна. Позволяет использовать градиентные методы оптимизации.
- Робастность: должна уметь работать с выбросами и не зависеть от небольшого числа экстремальных значений.
- Гладкость: должна иметь непрерывный градиент и не иметь резких переходов или скачков.
- Разреженность: должна стимулировать модель к получению разреженных результатов (уменьшение числа признаков).
- Монотонность: значение уменьшается по мере приближения прогнозируемого выхода к истинному.

Чему следовать выбирая функцию потерь

- Чувство прекрасного;
- Возможность применять математику:
 - Скорость вычисления;
 - Дифференцируемость (градиентные методы).
- Наличие интересных внутренних параметров;
- Возможность проверить осмысленность промежуточных результатов.

NB: Чем больше мы в области, тем больше знания мы перенесем в таргет.

Проблемы в построении

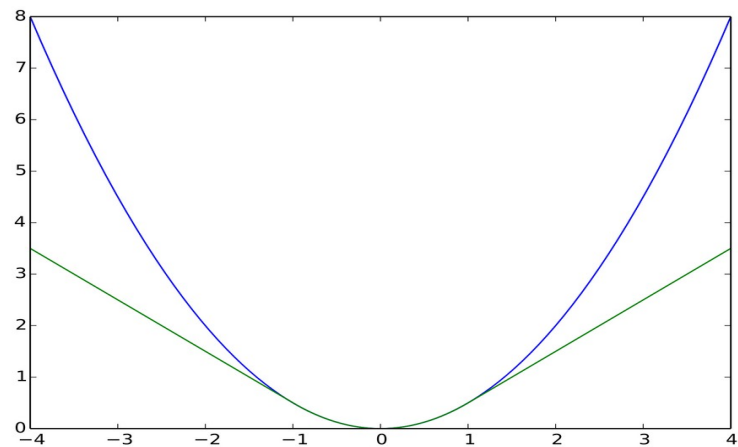
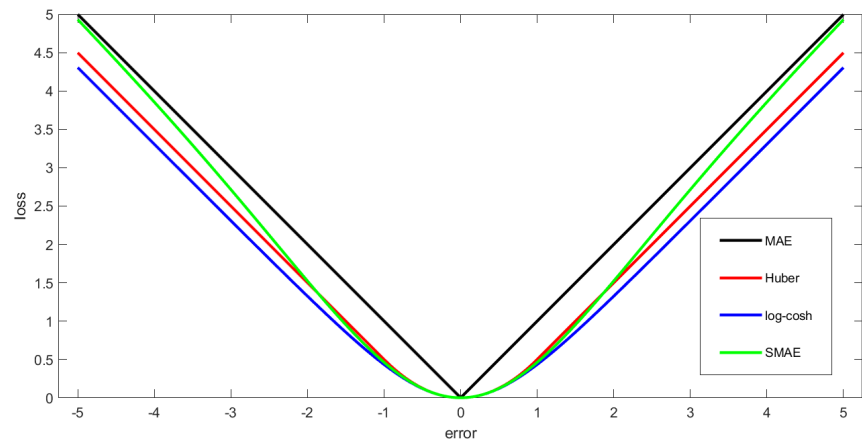
Что может быть “не так” в очевидном решении:

- Функция может быть неудобна для оптимизации;
- сложно гарантировать несмещённость по параметрам оптимизации;
- сложно собирать данные в терминах функции потерь

Стандартные функции потерь

<https://arxiv.org/pdf/2307.02694>

Регрессия



Регрессия

- MSE
 - дифференцируемая
 - выпуклая
 - чувствительна к выбросам
 - зависит от масштаба
- MAE
 - недифференцируемая
 - выпуклая
 - менее чувствительная к выбросам
 - зависит от масштаба
- Huber loss

Классификация

- Binary Cross Entropy (log loss)

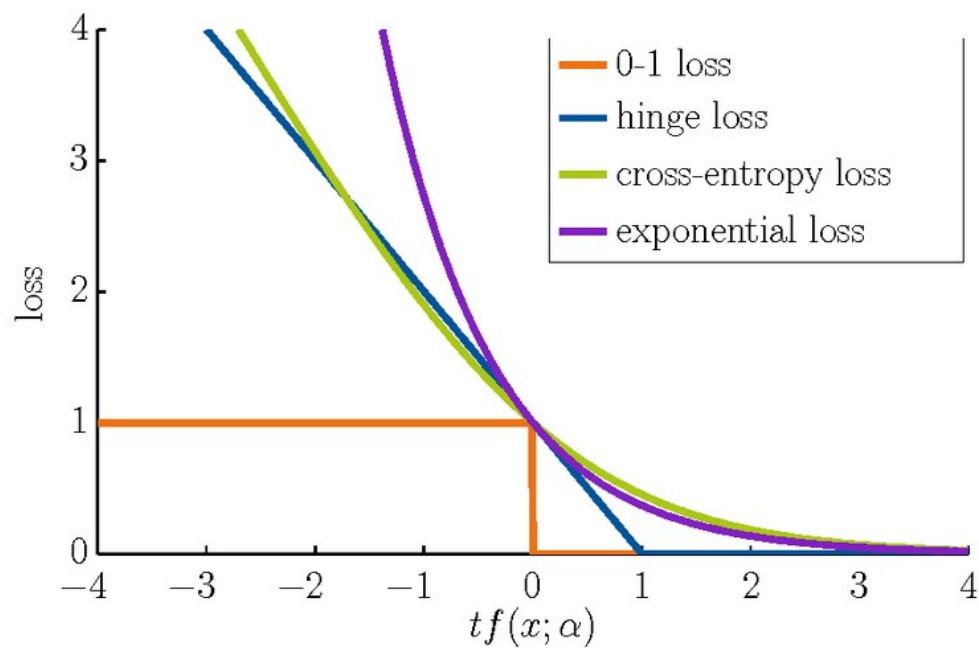
$$L(y, p) = -(y \log(p) + (1 - y) \log(1 - p))$$

- Cross Entropy loss
- Hinge loss

$$L(y, p) = \max(0, 1 - y \cdot p)$$

- Focal loss

Классификация



Классификация

- VSE
 - Дифференцируемая
 - Вероятностная интерпретация
 - Штраф зависит от уверенности предсказания
 - Чувствительна к несбалансированности классов
- Hinge loss
 - Выход +1 или -1
 - Штрафует, даже если ответ верный, но слишком близко к границе

Регуляризация

$$\hat{y} = f(x, w)$$

$$L(w) + R(w) \rightarrow \min$$

Метрики качества

Задача регрессии

- MSE
- RMSE
- MAE
- MBE
-

Задача регрессии

- MAPE
$$\text{MAPE} = 100 \frac{1}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right|$$
 - Плохо для близких к нулю значений
 - Больше штрафует за отрицательные ошибки
 - При недопрогнозировании ошибка не более 100%, при перепрогнозировании не ограничена
- SMAPE
$$\text{SMAPE} = \frac{100}{2} \sum_{t=1}^n \frac{|F_t - A_t|}{(|A_t| + |F_t|)/2}$$
 - По-разному штрафует за перепрогнозирование и недопрогнозирование
 - Плохо для близких к нулю значений

Задача регрессии

Коэффициент детерминации R^2 — это доля дисперсии зависимой переменной, объясняемая рассматриваемой моделью зависимости

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

- Хорош для линейных моделей
- Может быть отрицательным (модель хуже среднего целевой переменной)
- Чувствителен к выбросам
- Значение увеличивается (не уменьшается) от добавления в модель новых переменных, даже если эти переменные никакого отношения к объясняемой переменной не имеют.

Бинарная классификация

- Confusion matrix
- Accuracy
- Precision
- Recall
- F-measure
- ROC-AUC

Confusion matrix

Confusion Matrix

	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True Positives (TPs)	False Positives (FPs)
Predicted Negative (0)	False Negatives (FNs)	True Negatives (TNs)

Бинарная классификация

- Accuracy – отношение числа правильно классифицированных примеров к общему числу

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

- Precision

$$Precision = \frac{TP}{TP + FP}$$

- Recall

$$Recall = \frac{TP}{TP + FN}$$

Бинарная классификация

F-measure (F-score, F1) -- среднее геометрическое между precision и recall.
ближе к меньшему из precision, recall

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Бинарная классификация

ROC-AUC применяется для оценки «вероятностной» классификации. Оценивает качество ранжирования объектов по вероятности принадлежности к целевому классу

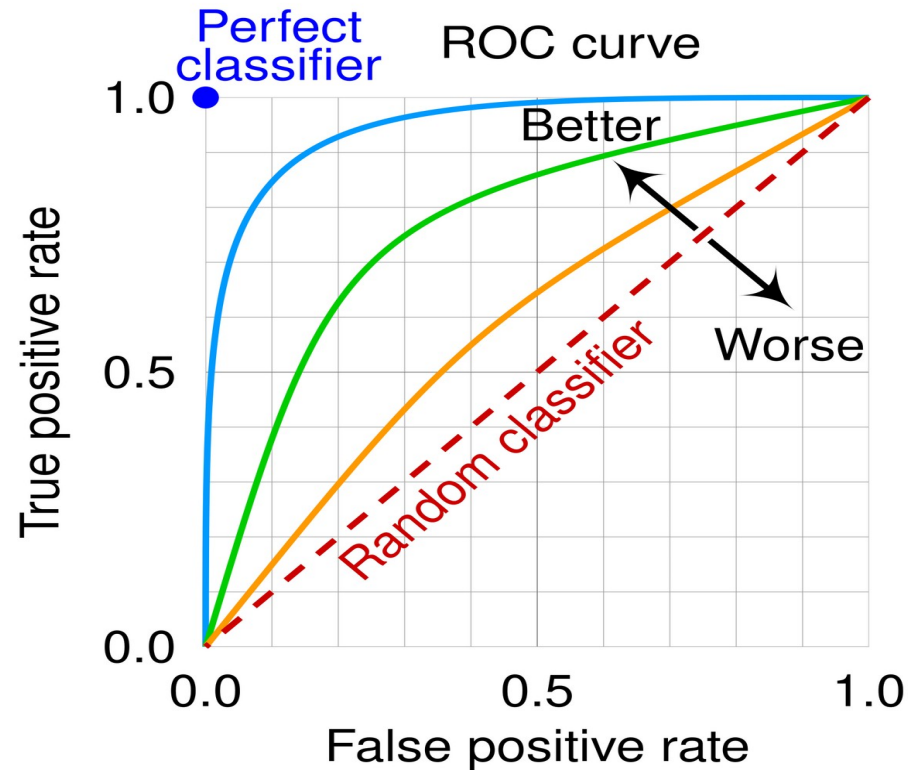
AUC – площадь под кривой

- Sensitivity (true positive rate) – отражает долю положительных результатов, которые правильно идентифицированы как таковые.
- Specificity (true negative rate) отражает долю отрицательных результатов, которые правильно идентифицированы как таковые.

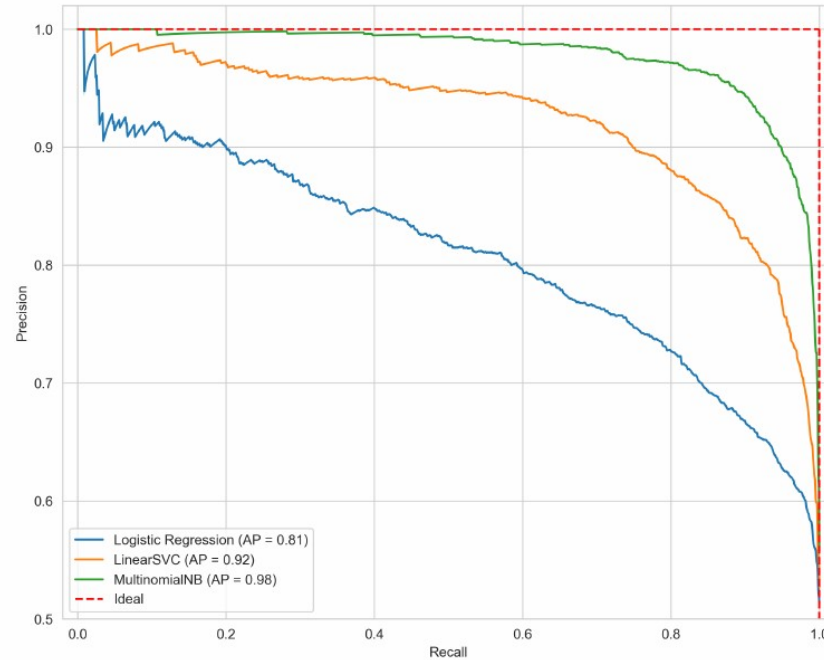
$$TPR = \frac{\textit{True positives}}{\textit{True positives} + \textit{False negatives}}$$

$$FPR = \frac{\textit{False positives}}{\textit{False positives} + \textit{True negatives}}$$

Бинарная классификация



Precision-Recall curve

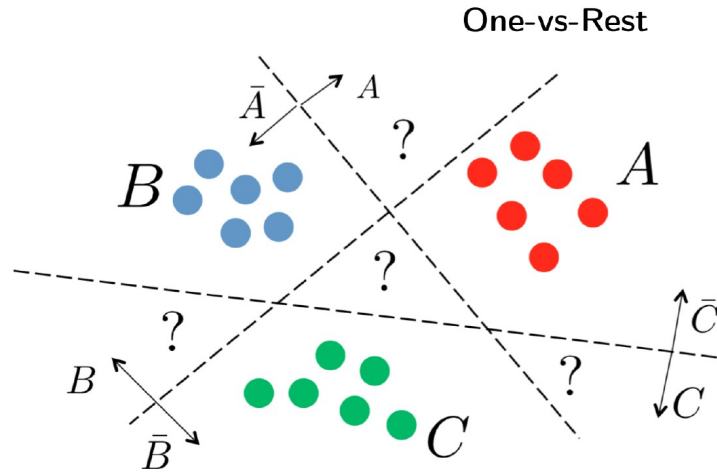


EER – equal error rate

Многоклассовая классификация

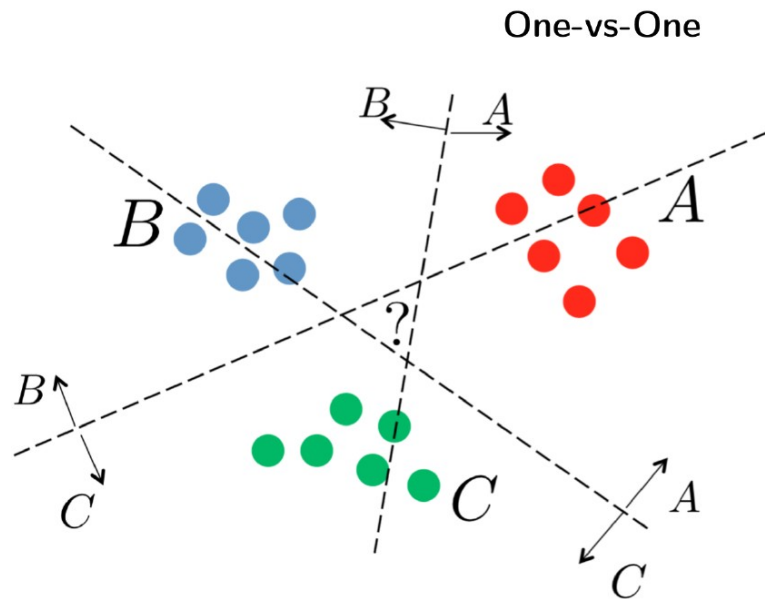
- Сведение к двухклассовому случаю
- Обобщение алгоритмов на многоклассовый случай

Сведение к двухклассовому случаю



K бинарных классификаторов

Сведение к двухклассовому случаю



$\frac{K(K-1)}{2}$ бинарных классификаторов

Многоклассовая классификация

- Confusion matrix
- Accuracy
- Recall и precision
 - Macro-average: независимо для каждого класса и усредняем
 - Micro-average: считаем сумму TP, FP, FN по всем классам
- Recall@k, Precision@k

Оценка качества системы

- Какие метрики использовать?
- На каких данных измерять?
- Метрики всей системы или отдельных частей?

На подумать

- Multilabel classification
- Sound events detection