

Цифровая обработка изображения

11. Функции потерь в задачах компьютерного
зрения

План занятия

- Функция потерь
- Классификация
- Сегментации
- Детекция
- Идентификация

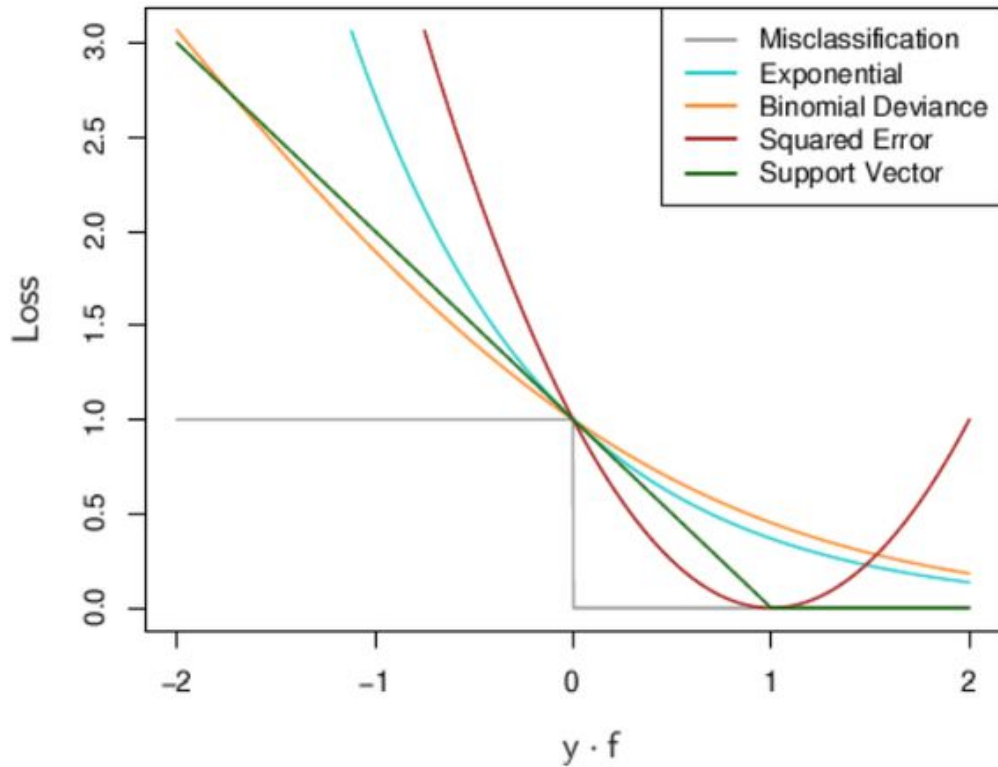
Функция потерь (Loss Function)

- оценка качества предсказания
- возвращает 0 в случае правильного ответа
- в противном случае возвращает значение пропорциональное величине ошибки (эмпирический риск)

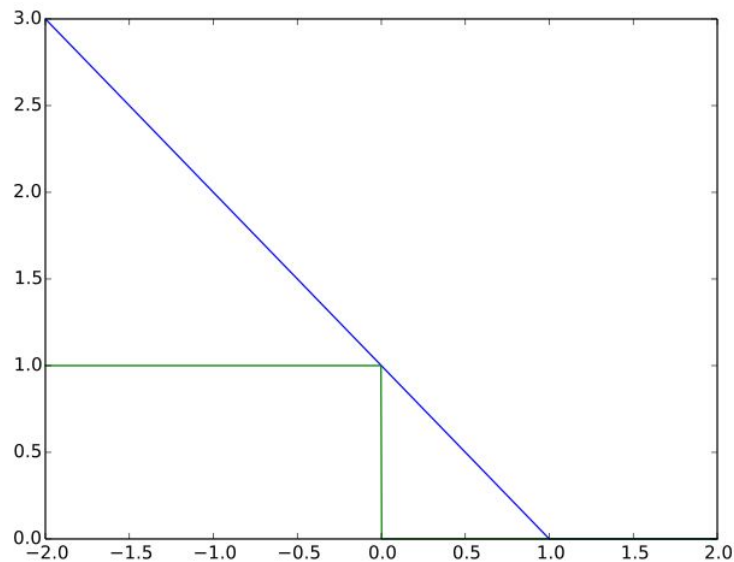
Функция потерь (Loss Function)

- требования к функции потерь:
 - непрерывность
 - дифференцируемость в каждой точке
 - выпуклость
- результирующее значение оценивается как среднее по всем примерам в выборке

Функция потерь: классификация

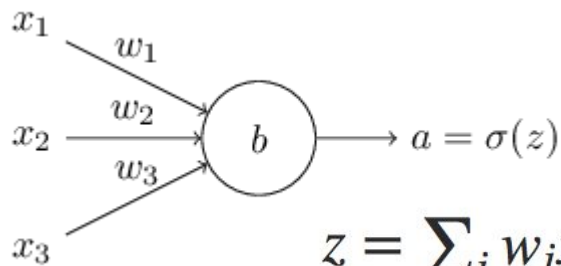


Классификация: Hinge Loss (SVM)



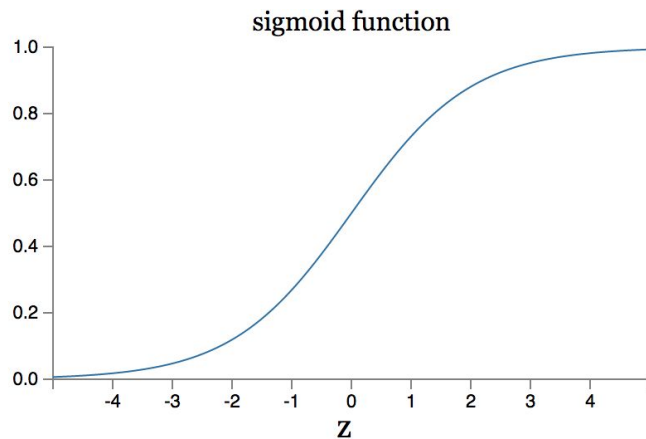
$$\ell(y) = \max(0, 1 - t \cdot y)$$

Классификация: Cross Entropy



$$z = \sum_j w_j x_j + b$$

$$\sigma(z) = 1/(1 + e^{-z})$$



$$C = -\frac{1}{n} \sum_x [y \ln a + (1 - y) \ln(1 - a)]$$

Классификация

- Hinge Loss
 - оптимизирует положение разделяющей гиперплоскости
- Cross Entropy
 - оценивает вероятность

Сегментация

- задаем количество классов
- дополнительно выбираем класс фона
- каждый пиксель изображения - отдельный пример в выборке
- для пикселя предсказываем вероятность принадлежности классам

Сегментация

- оценка градиента по семплу точек
- взвешивание примеров:
 - классам с меньшей площадью присваиваем больший вес
 - присваиваем больший вес точкам на границах объектов

Детекция объектов на изображении

- функция потерь состоит из двух частей
 - классификация
 - локализация (положение и размер)
- дополнительный бинарный выход, предсказывающий наличие объекта
- важно сбалансировать различия масштабов

Детекция

- классификационная функция потерь
 - Cross Entropy
- наличие объекта
 - Log Loss
- локализационная функция потерь
 - Регрессия

Детекция

i = anchor index in minibatch

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*).$$

Diagram annotations:

- Blue arrows point from $\{p_i\}$ and $\{t_i\}$ to the text: "Coordinates of the predicted bounding box for anchor i ".
- A blue arrow points from the text: "Predicted probability of being an object for anchor i " to p_i .
- A purple arrow points from "Log loss" to L_{cls} .
- A red arrow points from "Ground truth objectness label" to p_i^* .
- A purple arrow points from "Smooth L1 loss" to L_{reg} .
- A red arrow points from "True box coordinates" to t_i^* .
- A red circle highlights λ , with a line pointing to the text: "In practice $\lambda = 10$, so that both terms are roughly equally balanced".

N_{cls} = Number of anchors in minibatch (~ 256)

N_{reg} = Number of anchor locations (~ 2400)

Детекция: локализация

$$L_{\text{loc}}(t^u, v) = \sum_{i \in \{x, y, w, h\}} \text{smooth}_{L_1}(t_i^u - v_i)$$

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise} \end{cases}$$

Идентификация: Triplet Loss

- векторное представление изображения
- изображения одного объекта должны иметь одинаковое представление
- разные объекты - должны различаться

Идентификация: Triplet Loss

$$\sum_i^N \left[\|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha \right]_+$$

Полезные материалы

- [Loss functions for classification](#)
- [On Loss Functions for Deep Neural Networks in Classification](#)
- [FaceNet: A Unified Embedding for Face Recognition and Clustering](#)
- [In Defense of the Triplet Loss for Person Re-Identification](#)