

Week 3 Practice

CSCI 567 Machine Learning

Fall 2021

Instructor: Haipeng Luo

1. MULTIPLE-CHOICE QUESTIONS: One or more correct choice(s) for each question.

1.1. Which of the following surrogate losses is not an upper bound of the 0-1 loss?

- (a) exponential loss: $\exp(-z)$
- (b) hinge loss: $\max\{0, 1 - z\}$
- (c) perceptron loss: $\max\{0, -z\}$
- (d) logistic loss: $\ln(1 + \exp(-z))$

Ans: c, d. Note that here the logistic loss is using e as the base, instead of 2.

1.2. The perceptron algorithm makes an update $\mathbf{w}' \leftarrow \mathbf{w} + \eta y_n \mathbf{x}_n$ with $\eta = 1$ when \mathbf{w} misclassifies \mathbf{x}_n . Using which of the following different values for η will make sure \mathbf{w}' classifies \mathbf{x}_n correctly?

- (a) $\eta > \frac{y(\mathbf{w}^T \mathbf{x}_n)}{\|\mathbf{x}_n\|_2^2}$
- (b) $\eta < \frac{-y(\mathbf{w}^T \mathbf{x}_n)}{\|\mathbf{x}_n\|_2^2 + 1}$
- (c) $\eta < \frac{-y(\mathbf{w}^T \mathbf{x}_n)}{\|\mathbf{x}_n\|_2^2}$
- (d) $\eta > \frac{-y(\mathbf{w}^T \mathbf{x}_n)}{\|\mathbf{x}_n\|_2^2}$

Ans: d. Solve $y_n \mathbf{w}'^T \mathbf{x}_n = y_n (\mathbf{w} + \eta y_n \mathbf{x}_n)^T \mathbf{x}_n > 0$ for η .

1.3. Which of the following is true?

- (a) Normalizing the output \mathbf{w} of the perceptron algorithm so that $\|\mathbf{w}\|_2 = 1$ changes its test error.
- (b) Normalizing the output \mathbf{w} of the perceptron algorithm so that $\|\mathbf{w}\|_1 = 1$ changes its

test error.

(c) When the data is linearly separable, logistic loss (without regularization) does not admit a minimizer.

(d) Minimizing 0-1 loss is generally NP-hard.

Ans: c, d. For c, note that when the data is separable, one can find \mathbf{w} such that $y_n \mathbf{w}^T \mathbf{x}_n \geq 0$ for all n . Scaling this \mathbf{w} up will always lead to smaller logistic loss $\sum_{n=1} \ln(1 + \exp(-y_n \mathbf{w}^T \mathbf{x}_n))$ and thus the function does not admit a minimizer.

1.4. Which of the following statement is correct for function $f(\mathbf{w}) = w_1 w_2$?

(a) $(0, 0)$ is the only stationary point.

(b) $(0, 0)$ is a local minimizer.

(c) $(0, 0)$ is a local maximizer.

(d) $(0, 0)$ is a saddle point.

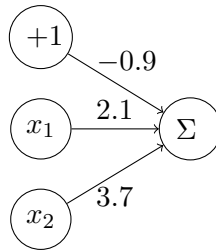
Ans: a, d. The gradient is $\nabla f(\mathbf{w}) = (w_2, w_1)$, so the only stationary point is $(0, 0)$. To see why it is neither a local minimizer nor a local maximizer, simply consider the direction $w_1 = -w_2$ and $w_1 = w_2$ respectively.

2. Perceptron

Consider the following training dataset:

\mathbf{x}	y
(0, 0)	-1
(0, 1)	-1
(1, 0)	-1
(1, 1)	1

and a perceptron with weights $(w_0, w_1, w_2) = \{-0.9, 2.1, 3.7\}$



2.1. What is the accuracy of the perceptron on the training data?

SOLUTION:

\mathbf{x}	y	\hat{y}	$\mathbb{I}(y = \hat{y})$
(0, 0)	-1	$\text{sgn}(-0.9) = -1$	Y
(0, 1)	-1	$\text{sgn}(-0.9 + 3.7) = 1$	N
(1, 0)	-1	$\text{sgn}(-0.9 + 2.1) = 1$	N
(1, 1)	1	$\text{sgn}(-0.9 + 2.1 + 3.7) = 1$	Y

Out of four predictions, two are correct. The accuracy is hence 50%.

- 2.2.** Select $\mathbf{x} = (1, 0)$ and $y = -1$. Use the perceptron training rule with $\eta = 1$ to train the perceptron for one iteration. What are the weights after this iteration?

For the given $\mathbf{x} = (1, 0)$ the classifier makes a mistake ($\hat{y} = 1$). We need to update the weights following the perceptron rule the new weights are given by

$$\begin{aligned} w_{k+1} &\leftarrow w_k + \eta y \mathbf{x} \\ &\leftarrow (-0.9, 2.1, 3.7) + (1)(-1)(1, 1, 0) \\ &\leftarrow (-1.9, 1.1, 3.7) \end{aligned}$$

- 2.3.** What is the accuracy of the perceptron on the training data after this iteration? Does the accuracy improve?

SOLUTION:

\mathbf{x}	y	\hat{y}	$\mathbb{I}(y = \hat{y})$
(0, 0)	-1	$sgn(-1.9) = -1$	Y
(0, 1)	-1	$sgn(-1.9 + 3.7) = 1$	N
(1, 0)	-1	$sgn(-1.9 + 1.1) = -1$	Y
(1, 1)	1	$sgn(-1.9 + 1.1 + 3.7) = 1$	Y

With the new weights, three out of four are correct, hence accuracy increased to 75%.

3. Maximum Likelihood Estimation

A random sample set X_1, X_2, \dots, X_n of size n is taken from a Poisson distribution with a mean of $\lambda > 0$. As a reminder, a Poisson distribution is a discrete probability distribution over the natural numbers, with the following probability mass function

$$P(X = x) = \frac{\lambda^x \cdot e^{-\lambda}}{x!}, \quad \forall x \in \{0, 1, 2, \dots, \}$$

3.1. Find the log likelihood of the data; call it $l(\lambda)$. You may use any log base you want.

$$\begin{aligned} l(\lambda) &= \log \prod_i P(X = X_i) \\ &= \log \prod_i \frac{\lambda^{X_i} e^{-\lambda}}{X_i!} \\ &= \sum_{i=1}^n [X_i \log \lambda - \lambda - \log(X_i!)] \\ &= \log \lambda \cdot \sum_{i=1}^n X_i - n\lambda - \sum_{i=1}^n \log(X_i!) \end{aligned}$$

3.2. Find the maximum likelihood estimator for λ .

$$\begin{aligned} l'(\lambda) &= \frac{1}{\lambda} \sum_{i=1}^n X_i - n = 0 \\ \lambda &= \frac{1}{n} \sum_{i=1}^n X_i \end{aligned}$$

So our maximum likelihood estimator $\hat{\lambda}$ is the average value.