

Exercises - Solutions

§6.2,6.3: Logistic Regression

Problem 1 Using $\sigma(t) = 1/(1 + e^{-t})$: $\sigma(-2) = \frac{1}{1+e^2} \approx 0.1192$. $\sigma(0) = 0.5$. $\sigma(2) = \frac{1}{1+e^{-2}} \approx 0.8808$. Then $1 - \sigma(2) \approx 1 - 0.8808 = 0.1192 = \sigma(-2)$.

Problem 2 $f(x) = -4 + 1.5x$. (a) $P(+1 \mid x) = 0.5$ when $f(x) = 0$. Thus $-4 + 1.5x = 0$, so $x = \frac{8}{3} \approx 2.667$. (b) At $x = 2$: $f = -4 + 3 = -1$. Therefore $P(+1 \mid 2) = \sigma(-1) \approx 0.2689$. (c) At $x = 4$: $f = -4 + 6 = 2$. Therefore $P(+1 \mid 4) = \sigma(2) \approx 0.8808$.

Problem 3 A one-unit increase in x_1 changes the log-odds by $\log 3$. Therefore the odds are multiplied by $e^{\log 3} = 3$. A two-unit increase changes the log-odds by $2 \log 3$. Thus the odds are multiplied by $e^{2 \log 3} = 3^2 = 9$.

Problem 4 (a) $y = +1, f = 2$. Margin: $m = 2$. Loss: $\log(1 + e^{-2}) \approx 0.1269$. Correctly classified. (b) $y = +1, f = -2$. Margin: $m = -2$. Loss: $\log(1 + e^2) \approx 2.1269$. Misclassified. (c) $y = -1, f = -2$. Margin: $m = 2$. Loss: 0.1269 . Correctly classified. (d) $y = -1, f = 0$. Margin: $m = 0$. Loss: $\log 2 \approx 0.6931$. The point is exactly on the decision boundary.

Problem 5 If $y = +1$, then $\sigma(yf) = \sigma(f)$, which is $P(Y = +1 \mid \mathbf{x})$. If $y = -1$, then $\sigma(yf) = \sigma(-f) = 1 - \sigma(f)$, which is $P(Y = -1 \mid \mathbf{x})$. Thus $P(Y = y \mid \mathbf{x}) = \sigma(yf)$. Now $-\log P(Y = y \mid \mathbf{x}) = -\log \sigma(yf) = -\log\left(\frac{1}{1+e^{-yf}}\right) = \log(1 + e^{-yf})$.

Problem 6 $\ell(m) = \log(1 + e^{-m})$. Differentiate: $\ell'(m) = \frac{-e^{-m}}{1+e^{-m}}$. Multiply numerator and denominator by e^m : $\ell'(m) = -\frac{1}{1+e^m}$. But $\sigma(-m) = \frac{1}{1+e^m}$. Therefore $\ell'(m) = -\sigma(-m)$. This is negative for every finite m . So increasing the signed margin lowers the loss, which is exactly what we want: correct and more confident classifications receive smaller loss.

Problem 7

```
scores = np.array([-3., -1., 0., 0.5, 2.])

probabilities = sigmoid(scores)

from_probability = np.where(
    probabilities >= 0.5, 1, -1
)

from_score = np.where(
    scores >= 0, 1, -1
)

print(probabilities)
print(from_probability)
print(from_score)
```

Probabilities are approximately (0.0474, 0.2689, 0.5, 0.6225, 0.8808). Both decision rules give

$(-1, -1, +1, +1, +1)$ using the stated convention that a tie at zero is assigned to $+1$.

Problem 8 For $f_1(x) = x - 3.5$, the margins are 2.5, 1.5, 0.5, 0.5, 1.5, 2.5. The average logistic loss is approximately 0.251. For $f_2(x) = 2(x - 3.5)$, the margins are doubled: 5, 3, 1, 1, 3, 5. The average logistic loss is approximately 0.123. Thus f_2 has the smaller logistic loss. However both scores have the same sign at every training observation. So both have classification error rate 0. This illustrates that logistic loss continues to reward larger correct margins even after the classification error has reached zero. For perfectly separable data, this is why unregularized logistic coefficients can continue growing.