

Exercises - Solutions

§6.2,6.7: Cross-Entropy

Problem 1 (a) $y = 1, p = 0.8$: $\ell = -\log(0.8) \approx 0.2231$. Correct prediction. (b) $y = 1, p = 0.2$: $\ell = -\log(0.2) \approx 1.6094$. Incorrect prediction. (c) $y = 0, p = 0.8$: $\ell = -\log(1 - 0.8) = -\log(0.2) \approx 1.6094$. Incorrect prediction. (d) $y = 0, p = 0.2$: $\ell = -\log(0.8) \approx 0.2231$. Correct prediction.

Problem 2 $C(p) = -[0.7 \log p + 0.3 \log(1 - p)]$. Differentiate: $C'(p) = -\frac{0.7}{p} + \frac{0.3}{1-p}$. Combine: $C'(p) = \frac{p-0.7}{p(1-p)}$. Thus $C'(p) = 0$ exactly when $p = 0.7$. Also $C''(p) = \frac{0.7}{p^2} + \frac{0.3}{(1-p)^2} > 0$ for $0 < p < 1$. Therefore $p = 0.7$ is the unique minimizer.

Problem 3 First, $\frac{\partial \ell}{\partial p} = -\frac{y}{p} + \frac{1-y}{1-p} = \frac{p-y}{p(1-p)}$. For $p = \sigma(f)$: $\frac{dp}{df} = p(1-p)$. By the chain rule: $\frac{\partial \ell}{\partial f} = \frac{p-y}{p(1-p)} p(1-p) = p - y$.

Problem 4 $p - y = 0.25 - 1 = -0.75$. Therefore $(p - y)\tilde{\mathbf{x}} = -0.75 \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} = \begin{bmatrix} -0.75 \\ -1.50 \\ 0.75 \end{bmatrix}$.

Problem 5 At $\mathbf{w} = \mathbf{0}$, all scores are zero. Thus $\mathbf{p} = \begin{bmatrix} 0.5 \\ 0.5 \\ 0.5 \end{bmatrix}$. Then $\mathbf{p} - \mathbf{y} = \begin{bmatrix} 0.5 \\ -0.5 \\ -0.5 \end{bmatrix}$. Therefore $\nabla g = \frac{1}{3} X^T (\mathbf{p} - \mathbf{y}) = \begin{bmatrix} -1/6 \\ -1/3 \end{bmatrix}$. With $\alpha = 0.3$: $\mathbf{w}_1 = \mathbf{w}_0 - 0.3 \nabla g = -0.3 \begin{bmatrix} -1/6 \\ -1/3 \end{bmatrix} = \begin{bmatrix} 0.05 \\ 0.10 \end{bmatrix}$.

Problem 6 Because $p = \sigma(f) = \frac{e^f}{1+e^f}$, we have $1 - p = \frac{1}{1+e^f}$. Then $-[y \log p + (1 - y) \log(1 - p)] = -y[f - \log(1 + e^f)] + (1 - y) \log(1 + e^f) = \log(1 + e^f) - yf$. Now let $s = 2y - 1$. If $y = 1$, then $s = 1$ and $\ell = \log(1 + e^{-f}) = \log(1 + e^{-sf})$. If $y = 0$, then $s = -1$ and $\ell = \log(1 + e^f) = \log(1 + e^{-sf})$. Thus the two forms are identical.

Problem 7 For observed class $y = 1$: at $p = 0.51$: $-\log(0.51) \approx 0.6733$. At $p = 0.80$: $-\log(0.80) \approx 0.2231$. At $p = 0.99$: $-\log(0.99) \approx 0.0101$. All three predictions are classified correctly using threshold 0.5. Therefore accuracy treats them identically. Cross-entropy rewards the model that assigns larger probability to the observed class.

Problem 8

```
def cross_entropy_from_scores(w, X, y):
    scores = X @ w
```

```

losses = np.logaddexp(0, scores) \
        - y*scores

return np.mean(losses)

def grad_cross_entropy(w, X, y):
    scores = X @ w
    p = sigmoid(scores)

    return X.T @ (p - y) / len(y)

```

For the given data: scores are -0.5 , 0.5 , 1.5 . Probabilities are approximately 0.3775 , 0.6225 , 0.8176 . The mean cross-entropy is approximately 0.3832 . The gradient is approximately $\begin{bmatrix} -0.0608 \\ -0.1866 \end{bmatrix}$.