Problem 1 — Linear Regression / SVD

(a) The empirical squared-loss risk is

$$\hat{R}(w) = \frac{1}{2n} \sum_{i=1}^{d} \sum_{j=1}^{n_i} (w_i - y_{ij})^2.$$

Differentiating coordinate-wise and setting to zero gives

$$w_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}.$$

(b) With $X = \sum_{t=1}^{r} s_t u_t v_t^\top$ and $y \in \text{span}\{u_1, \ldots, u_r\}$, let

$$w = \sum_{t=1}^{r} \frac{\beta_t}{s_t} v_t, \qquad \text{where } y = \sum_{t=1}^{r} \beta_t u_t.$$

Then

$$Xw = y,$$

so the empirical risk is zero.

(c) The nonzero eigenvalues of $X^\top X$ are $s_1^2, \ldots, s_r^2$. If the rows of X span $\mathbb{R}^d$, then $\text{rank}(X) = d$ and $X^\top X$ is invertible. Conversely, if $X^\top X$ is invertible, then $\text{rank}(X) = d$, hence the rows span $\mathbb{R}^d$.

(d) Example:

$$X = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad X^\top X = I_2 \text{ (invertible)}, \qquad XX^\top = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \text{ (not invertible).}$$

Problem 2 — Linear Regression

(a), (b) [skiped: code] **(c)**

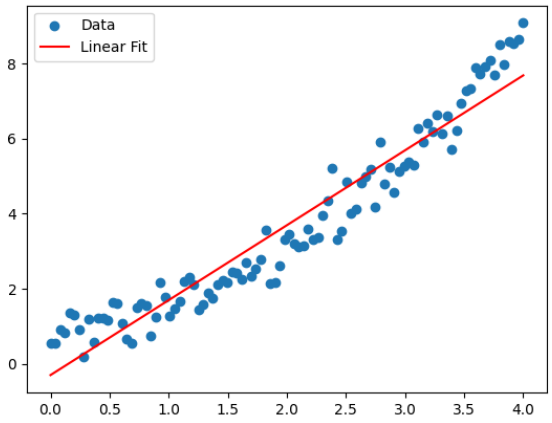

Figure 1: Linear fit and data (Problem 2(c))

Problem 3 — Polynomial Regression

(a) For $x = (x_1, x_2, x_3)$,

$$\phi(x) = \left[1,\ x_1,\ x_2,\ x_3,\ x_1^2,\ x_1x_2,\ x_1x_3,\ x_2^2,\ x_2x_3,\ x_3^2\right]^\top.$$

(b) to (e) [skiped: code]

Problem 4 — Logistic Regression

(a)

$$\hat{R}_{\log}(w) = \frac{1}{n}\sum_{i=1}^{n} \ln\bigl(1 + \exp(-y_i w^\top x_i)\bigr),$$

$$\nabla_w \hat{R}_{\log}(w) = -\frac{1}{n}\sum_{i=1}^{n} y_i x_i\, \sigma(-y_i w^\top x_i),$$

$$w' = w + \frac{\eta}{n}\sum_{i=1}^{n} y_i x_i\, \sigma(-y_i w^\top x_i).$$

(b), (c) [skipped: code]

Problem 5 — N-Gram Next Token Prediction (Cross-Entropy)

(a) For sample (x_i, y_i) with one-hot target e_{y_i},

$$\nabla_W \ell_i(W) = x_i\big(p(\cdot|x_i) - e_{y_i}\big)^\top,$$
$$\nabla_W \hat{R}_{\mathrm{CE}}(W) = \frac{1}{n}\sum_{i=1}^{n} x_i\big(p(\cdot|x_i) - e_{y_i}\big)^\top,$$
$$W' = W - \frac{\eta}{n}\sum_{i=1}^{n} x_i\big(p(\cdot|x_i) - e_{y_i}\big)^\top.$$

(b) to (e) [skipped: code]

Problem 6 — LLM Use, Collaboration, and Other Sources

1. An AI assistant was used in Problems 1(a–d), 4(a), and 5(a). Final answers were edited and wrapped up manually on my ipad. Then, used AI to reformat into Latex.

2. The assignment handout was referenced for problem statements and notation.

3. No additional external sources or collaborators were used.