Partial solutions to practice problems for Exam 2

COMS 4771 Fall 2016

Problem 1

(a) $f''(\langle w, x \rangle) w w^\top \succeq 0$. This implies that g is convex.

(b) Yes. Note that h is differentiable but not (necessarily) twice differentiable. So we'll check the first-order condition for convexity. Pick any $x, x_0 \in \mathbb{R}^d$. Then

$$\nabla h(x_0) = \begin{cases} f'(\langle w, x_0 \rangle) w & \text{if } \langle w, x_0 \rangle \leq 0, \\ f'(0) w & \text{if } \langle w, x_0 \rangle > 0. \end{cases}$$

- If $\langle w, x_0 \rangle \leq 0$ and $\langle w, x \rangle \leq 0$, then

$$\begin{aligned} h(x_0) + \langle \nabla h(x_0), x - x_0 \rangle &= g(x_0) + f'(\langle w, x_0 \rangle) \langle w, x - x_0 \rangle \\ &= g(x_0) + \langle \nabla g(x_0), x - x_0 \rangle \\ &\leq g(x) \\ &= h(x) \end{aligned}$$

  where the inequality follows because g is convex.

- If $\langle w, x_0 \rangle \leq 0$ and $\langle w, x \rangle > 0$, then

$$\begin{aligned} h(x_0) + \langle \nabla h(x_0), x - x_0 \rangle &= g(x_0) + \langle \nabla g(x_0), 0 - x_0 \rangle + f'(\langle w, x_0 \rangle) \langle w, x \rangle \\ &\leq f'(\langle w, x_0 \rangle) \langle w, x \rangle \\ &\leq f'(0) \langle w, x \rangle \\ &= h(x) \end{aligned}$$

  where the first inequality follows because g is convex and $g(0) = f(0) = 0$, and the second inequality follows because f' is non-decreasing, $\langle w, x_0 \rangle \leq 0$, and $\langle w, x \rangle > 0$.

- If $\langle w, x_0 \rangle > 0$ and $\langle w, x \rangle \leq 0$, then

$$\begin{aligned} h(x_0) + \langle \nabla h(x_0), x - x_0 \rangle &= f'(0) \langle w, x \rangle \\ &= g(0) + \langle \nabla g(0), x - 0 \rangle \\ &\leq g(x) \\ &= h(x) \end{aligned}$$

  where the inequality follows because g is convex.

- If $\langle w, x_0 \rangle > 0$ and $\langle w, x \rangle > 0$, then

$$\begin{aligned} h(x_0) + \langle \nabla h(x_0), x - x_0 \rangle &= f'(0)\langle w, x_0 \rangle + f'(0)\langle w, x - x_0 \rangle \\ &= f'(0)\langle w, x \rangle \\ &= h(x)\,. \end{aligned}$$

This implies that h satisfies the first-order condition for convexity, so h is convex.

Problem 2

(a)

(b)

(c) The only difference from part (b) is that each iterate must be projected to the unit ball $\{w \in \mathbb{R}^d : \|w\|_2 \leq 1\}$. This can be done using the following procedure:

- Input: vector $w \in \mathbb{R}^d$.
- If $\|w\|_2 \leq 1$, return w.
- Else, return $w/\|w\|_2$.

Problem 3

(a) Yes
(b) No
(c) Yes

Problem 4

1. Function class $\mathcal{F}$ may not contain log-odds function.
2. Random sample may not be representative of P.
3. Algorithm may not find good solution to optimization problem.

Problem 5

$2\mu - 1$

Problem 6

(a) 0.5
(b) 0.5
(c) 0.55

Problem 7

Let $D := \mathrm{diag}(\|x_1\|_2^2, \ldots, \|x_n\|_2^2)$ and $b = (y_1, y_2, \ldots, y_n)$.

MLE is $(A^\top D^{-1} A)^{-1} A^\top D^{-1} b$.

Problem 8

Problem 9

(a) (Here, we drop the subscript i on the vectors a_i and b_i.) The (i,i)-th entry of the matrix $ab^\top$ is $a_i b_i$. So $\mathrm{tr}(ab^\top) = a_1 b_1 + \cdots + a_d b_d = \langle a, b \rangle$.
(b) Since $A^\top B = \sum_{i=1}^n a_i b_i^\top$ and the (i,i)-th entry of $AB^\top$ is $\langle a_i, b_i \rangle$, we have $\mathrm{tr}(A^\top B) = \sum_{i=1}^n \mathrm{tr}(a_i b_i^\top) = \sum_{i=1}^n \langle a_i, b_i \rangle = \mathrm{tr}(AB^\top)$.
(c) Let $\Sigma := \mathbb{E}[(X - \mathbb{E}(X))(X - \mathbb{E}(X))^\top] = \mathbb{E}[XX^\top]$ be the covariance matrix of the mean-zero random vector X. Then $\mathrm{tr}(\Sigma) = \mathrm{tr}(\mathbb{E}[XX^\top]) = \mathbb{E}[\mathrm{tr}(XX^\top)] = \mathbb{E}[\langle X, X \rangle] = \mathbb{E}[\|X\|_2^2]$. Moreover, writing the eigendecomposition of Σ as $\Sigma = V \Lambda V^\top$, we have $\mathrm{tr}(\Sigma) = \mathrm{tr}(V \Lambda V^\top) = \mathrm{tr}(V^\top V \Lambda) = \mathrm{tr}(\Lambda) = \lambda_1 + \lambda_2 + \cdots + \lambda_d$.
(d) $\lambda_1 + \lambda_2 + \cdots + \lambda_d = \mathbb{E}\|X - \mathbb{E}(X)\|_2^2$

Problem 10

(a) $(VS^2V^\top)w = VSU^\top y$.
(b) Multiply both sides of normal equations by $VS^{-2}V^\top$ to get

$$VV^\top w = VS^{-1}U^\top y.$$

(c) Any w that satisfies the normal equations can be written as $w = VV^\top w + (I - VV^\top)w = A^\dagger y + (I - VV^\top)w$. By Pythagorean theorem, $\|w\|_2^2 = \|VV^\top w\|_2^2 + \|(I - VV^\top)w\|_2^2 = \|A^\dagger y\|_2^2 + \|(I - VV^\top)w\|_2^2$, which is strictly larger than $\|A^\dagger y\|_2^2$ whenever $w \neq A^\dagger y$.

Problem 11

(a) Consider the $k = 1$ case. The average of rank one matrices $uu^\top$ and $vv^\top$ can have rank two (e.g., when u and v are linearly independent). So the set of rank one matrices is not convex.
(b) $k = 0$ and $k = \min\{m, n\}$.
(c) $\Omega = \{1, 2, \ldots, m\} \times \{1, 2, \ldots, n\}$.

Problem 12

$\langle a_i, b - \sum_{j \neq i} \hat{w}_j a_j \rangle / \|a_i\|_2^2$.

Time complexity: $O(nd)$.

Problem 13

No. The k-means objective from lecture can be written as

$$\sum_{i=1}^{k} \sum_{x \in C_i} \|x - c_i\|_2^2 \,,$$

but the objective from ESL is equal to

$$\sum_{i=1}^{k} |C_i| \sum_{x \in C_i} \|x - c_i\|_2^2 \,.$$

Problem 14

(a) Let $l(t) := \text{logistic}(t)$. Then

$$\frac{l(\langle \hat{\alpha}, x_i \rangle) \cdot l(\langle \hat{\beta}_1, x_i \rangle)^{y_i} \cdot l(-\langle \hat{\beta}_1, x_i \rangle)^{1-y_i}}{l(\langle \hat{\alpha}, x_i \rangle) \cdot l(\langle \hat{\beta}_1, x_i \rangle)^{y_i} \cdot l(-\langle \hat{\beta}_1, x_i \rangle)^{1-y_i} + l(-\langle \hat{\alpha}, x_i \rangle) \cdot l(\langle \hat{\beta}_0, x_i \rangle)^{y_i} \cdot l(-\langle \hat{\beta}_0, x_i \rangle)^{1-y_i}} \,.$$

(b) This is very similar to "mixtures of two linear regressions" example from lecture. In the M-step, instead of a weighted logistic regression problem and two weighted least squares problems in the M-step, we have now three weighted logistic regression problems.

Problem 15

E-step:

$$w_i \ := \ \frac{\hat{\pi}_i \prod_{j=1}^m \hat{r}_j^{x_{i,j}} (1-\hat{r}_j)^{1-x_{i,j}}}{\hat{\pi}_i \prod_{j=1}^m \hat{r}_j^{x_{i,j}} (1-\hat{r}_j)^{1-x_{i,j}} + (1-\hat{\pi}_i) \prod_{j=1}^m \hat{p}_j^{1-x_{i,j}} (1-\hat{p}_j)^{x_{i,j}}}$$

for all $i = 1, 2, \ldots, m$.

M-step:

$$\begin{aligned}
\hat{\pi}_i &:= w_i \quad \text{for all } i = 1, 2, \ldots, m\,, \\
\hat{p}_j &:= \frac{\sum_{i=1}^m (1-w_i)(1-x_{i,j})}{\sum_{i=1}^m (1-w_i)} \quad \text{for all } j = 1, 2, \ldots, n\,, \\
\hat{r}_j &:= \frac{\sum_{i=1}^m w_i x_{i,j}}{\sum_{i=1}^m w_i} \quad \text{for all } j = 1, 2, \ldots, n\,.
\end{aligned}$$