Q/A Assignment.

Solution 1.

(For a single point)

For logistic regression

loss function $= y \log \hat{y} + (1-y) \log(1-\hat{y})$

where y is the label and

$\hat{y}$ is the model's output using input vector x

also $\hat{y} = \sigma(W.x)$ where $w = [w_1, w_2, \ldots w_n]$ are the weights

& $x = [x_1 \ldots x_n]$ are features

Since logistic regression is trained on both cases we can assume loss for both models will be almost same.

Let $\omega = [w_1 \ldots w_n]$ & $\omega_{new} = [\omega_{new,1}, \ldots \omega_{new,n+1}]$

& $x = [x_1, \ldots x_n]$ & $x_+ = [x_1 \ldots x_n, x_n]$.

$$\sigma(wx)$$

so $y \log(\sigma(Wx)) + (1-y)\log(1-\hat{y}) = y\log(\sigma(w_{new}x_+))$
$+ (1-y)\log(1-\sigma(w_{new}x_+))$

Since this should be true for all points

so let $y = 1$ and $x_1 = x_2 = x_{n-1} = 0$

$\Rightarrow \sigma(wx) = \sigma(\omega_{new} x_+)$

$\Rightarrow w_n \cancel{x_n} = \omega_{new,n}\cancel{x_n} + \omega_{new,n+1}\cancel{x_n}$

$\Rightarrow \boxed{w_n = (\omega_{new,n} + \omega_{new,n+1})}$ & $\boxed{w_{new,i} = w_i \quad \forall i \in [1, n-1]}$

Q.2 / Sol-2.

Using pairwise comparision ~~to per wis~~ with z- test between each template and controll.

- Compairing A with E, CTR ~~of~~ E is significantly higher than A.
- B ~~has~~ slightly lower CTR than A.
- D has higher CTR than A.
- C has slighty higher CTA than A {although comparable}

So option $\boxed{C}$ is correct.

Sol 3.

Since g.d iteration =>

$$\omega_{t+1} = \omega_t - \alpha \nabla_\theta f(\theta, z)$$

where $f(\theta, z)$ is the output of loss

let the cost function be $l(\cdot,\cdot)$ ~~and~~ $O(\theta, x)$ be the output of model.

i.e

$$f(\theta, z) = l(y, O(\theta, x))$$

so

$$\nabla_\theta f(\theta, z) = l'(y, O(\theta, x)) \nabla_\theta O^*(\theta, x)$$

Because x is sparse so for $O(\theta, x)$ instead of n operations you will only need k operations

i.e. $x_1 w_1 + x_2 w_2 + \dots x_n w_n \approx O(k)$ instead of $O(n)$

For $\nabla_\theta O(\theta, x)$

since $\frac{\partial O(\theta, x)}{\partial w_i} = x_i$

so computation of vector $\nabla_\theta O(\theta, x)$ will again take $O(k)$ instead of $O(n)$.

~~$\therefore$ This so final increase will be of the order. $O($~~

So final computation cost will be $O(k^2)$.

(considering efficient sparse vector $\nabla_\theta O(\theta, x)$ ~~product~~ would be represented in $O(k)$ time.)