Name → Ambuj Mishra

Roll No. → 20216003

Q(1) → $\boxed{X_i = Y_i - f_0(x_i)}$;

this'll denote the deviation b/w the actual & estimated values.

So → let's take → $\boxed{Y = \phi(x) \cdot \mu + \epsilon}$

$\boxed{Y - \phi(x) \cdot \mu = \epsilon}$ error

Correct statements → {error will be Normally distributed.

(a). Correct → $X_i \sim N(0,1)$

(b). $E(X_i) = 0$; Correct. ⇒ bcz we ideally consider that deviation of error from mean is same on both sides. Therefore expected value = 0

(c). $Var(X_i) = 1$; Correct.

⇓

from part (a), since SD = 1

thus $Var(X_i) = 1$

Q(2) → $f(x) = (w_1 x_1 + w_2 x_2 + b)$; where $w_1 = w_2 = 0$ & $b = 0.5$

$\boxed{\text{Training RMSE} = 0.5}$

x_1	x_2	$Y_{act.}$	$Y_{est.}$
0	0	0	0.5
0	1	1	0.5
1	0	1	0.5
1	1	0	0.5

(b). high bias → error b/w estimated & ~~pre~~ actual value is very high.

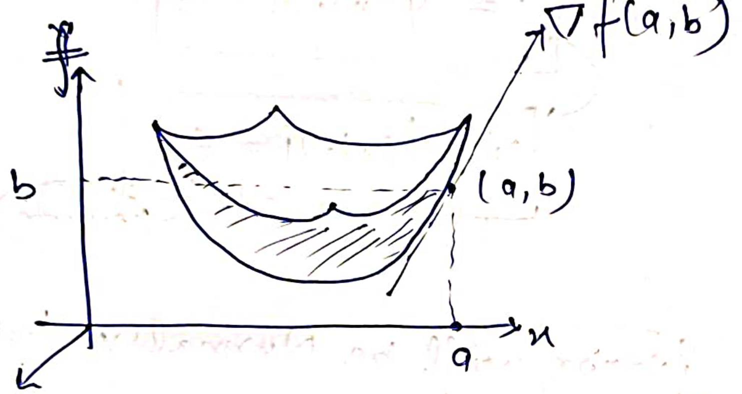
Correct.

(4.)

(b)

→ bias term is more dominating only in the estimated fn.

Q5. → $f: R^n \rightarrow R$



from the curve, we know that we want to calculate the estimation fn which minimises the loss of fn, So → after making the loss fn. convex, we want to find the minima of the curve.

In case of convex fn, minima always lies in the opposite dirn of gradient.

So, for pt. (a,b) → if gradient is in the dirn ⇒

$\nabla f(a,b)$ → then we always move in the opposite of gradients to reach minima.

Thus →

$$\boxed{w^{k+1} = w^k - \eta \cdot \nabla f(a,b)}$$

(b) → in case of Regularised least-square estimator, curve is more generalised (thus low variance).

But if the dataset is more noisy, then error increases. (thus high bias)

→ in case of training data having noises.

if the testing data has no noises, & it's generalised, then → low bias on testing.

answer → (c) or (d) depending upon existence of noise in test data. ⇒

if test data is not noisy, then ⇒ (d).

---

(8) → if we increase the number of training pts ⇒

So, in case of very large size of training data (>>> 1) →

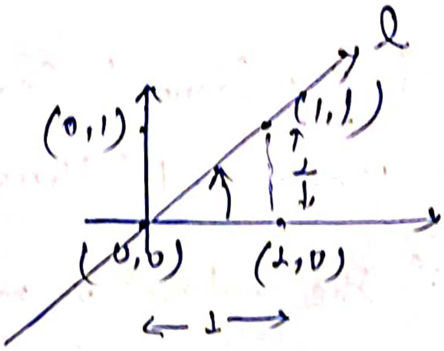
(a). The bias of model decreases → bcoz large training pts. will average out the effect of noisy datapoints, which will lower the bias.

(c). variance of model decreases → large no. of datapts. will reduce overfitting and will generalise the estimation model.

Thus → low variance.

(7).

x_1	x_2
1	0
0	1
1	0
1	1



variance will be maximum along the (l).

thus → $\cancel{\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)} \Rightarrow \cancel{\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)}$

$\tan\theta = 1 \Rightarrow \theta = 45^\circ$.

Therefore sol.^{n} ⇒ (0.5, 0.5).

⇒ option (a).

④ → to compute L1-Norm →

① first calculate gradient →

②. Main eq.ⁿ →

$$\Rightarrow \text{Min.}\ \frac{1}{\alpha}\lambda^T\lambda + \frac{1}{2}\text{Min.}\ \cancel{\Sigma(Y_i - \phi(x_i)\omega + b)}$$

$$\underbrace{\Sigma\{Y_i - (\phi(x_i)\cdot\omega + b)\}}_{H}$$

$$\Downarrow$$

if

$$H > 0 \Rightarrow (-x)$$

$$H < 0 \Rightarrow (+x)$$

③ → Then calc. using.

$$\boxed{\begin{bmatrix} \omega^{k+1} \\ b^{k+1} \end{bmatrix} = \begin{bmatrix} \omega^{k} \\ b^{k} \end{bmatrix} - \eta \cdot \begin{bmatrix} \nabla f_\omega \\ \nabla f_b \end{bmatrix}}$$

④. Do this for a tolerance