3.

(a) Show that solutions to

$\hat{a}(\lambda) = \underset{a}{\text{argmin}}\ \hat{R}(a) + \lambda \cdot P_\gamma(a)$ have nonzero values for all coefficients for the entire path indexed by $\lambda$, for some arbitrary $\lambda$, when

- $\hat{R}(a) = \frac{1}{N} \sum_{i=1}^{N} (y_i - \bar{a} \cdot \bar{x})^2$
- $P_\gamma(a) = \sum_{j=1}^{n} |a_j|^\gamma \quad (\gamma > 1)$

Since both $\hat{R}(a)$ & $P_\gamma(a)$ are convex and differentiable, we know the value of a that minimizes $h_{\lambda,\gamma}(a) = \hat{R}(a) + \lambda P_\gamma(a)$ occurs when $\frac{\partial h}{\partial a_j} = 0$, $j = 1 \ldots n$. We also know there is one global minimum, $a_j^*$, $j = 1 \ldots n$.

For a particular $a_j$, we can solve for $a_j^*$ by setting $\frac{\partial h}{\partial a_j} = 0$

$$\frac{\partial h}{\partial a_j} = \frac{\partial \hat{R}(a)}{\partial a_j} + \lambda \frac{\partial P_\gamma(a)}{\partial a_j}$$

$$\frac{\partial h}{\partial a_j} = -\frac{2}{N} \sum_{i=1}^{N} (y_i - \bar{a} \cdot \bar{x}) x_{ij} + \lambda \gamma (|a_j|)^{\gamma-2} (a_j)$$

This derivative is 0 when $a_j = a_j^*$.

$$0 = \frac{-2}{N} \sum_{i=1}^{N} (y_i - \bar{a} \cdot \bar{x}) x_{ij} + \lambda \gamma (|a_j^*|)^{\gamma-2} (a_j^*)$$

Now, suppose $a_j^* = 0$. This will only be the argmin when

$$0 = -\frac{2}{N} \sum_{i=1}^{N} (y_i - \bar{a}\cdot\bar{x})\, x_{ij}$$

This will occur with probability $\sim 0 \rightarrow$ so we can assume that in general, $a_j^* \neq 0 \;\forall j = 1 \ldots n$. $\square$

(b) Show that with probability $> 0$, the quantity $a_j^*$ that minimizes $\underset{a}{\text{argmin}}\, \hat{R}(a) + \lambda P_\gamma(a)$ can be 0 for an arbitrary $\lambda$, when $P_\gamma(a) = \sum_{j=1}^{n} (\gamma-1)\, a_j^2/2 + (2-\gamma)|a_j|$ $(1 \leq \gamma < 2)$.

Since $P_\gamma(a)$ is not differentiable, we cannot take the gradient and solve for when it is 0. So instead we will show that it is possible for $a_j^* = 0$ by arguing that if we are in a situation where $a_j^* = 0$, then moving in the direction $e_j$ should increase $\hat{R}(a) + \lambda P_\gamma(a)$, and moving in the direction $-e_j$ will also increase the risk. So:

WLOG, assume $j = 1$.

$$D_{e_1} = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon}\left[\hat{R}(a+\varepsilon e_1) + \lambda P(a+\varepsilon e_1) - \hat{R}(a) - \lambda P(a)\right]$$

$$= \langle \nabla R, e_1 \rangle + \lim_{\varepsilon \to 0} \frac{1}{\varepsilon}\left[\lambda P(a+\varepsilon e_1) - \lambda P(a)\right]$$

$$= -\frac{2}{N} \sum_{i=1}^{N} (y - \bar{x}\bar{a})\, x_1 + \lambda \lim_{\varepsilon \to 0} \frac{1}{\varepsilon}\Big[(\gamma-1)(a_1+\varepsilon)^2/2 +$$

$$(2-\gamma)|a_1+\varepsilon| - (\gamma-1)a_1^2/2 - (2-\gamma)|a_1|\Big]$$

$$= \langle \nabla R, e_1 \rangle + \lim_{\varepsilon \to 0} \frac{1}{\varepsilon}\left[\frac{(\gamma-1)}{2}\left[a_1^2 + 2a_1\varepsilon + \varepsilon^2 - a_1^2\right] + (2-\gamma)\,|a_1+\varepsilon| - |a_1|\right.$$

$$= \langle \nabla R, e_1 \rangle + \lambda \lim_{\varepsilon \to 0} \left[ \frac{(y-1)}{2} [2a_1 + \varepsilon] + (2-y) \left[ |a_1/\varepsilon + 1| - |a_1/\varepsilon| \right] \right]$$

$$= \langle \nabla R, e_1 \rangle + \lambda \frac{(y-1)}{2} [2a_1] + (2-y)$$

$$D_{e_1} = \frac{-2}{N} \sum_{i=1}^{N} (y - \bar{x}\bar{a}) x_1 + \lambda \left[ (y-1) a_1 + (2-y) \right]$$

This quantity is greater than 0 for $a_1 = 0$ when

$$\frac{2}{N} \sum_{i=1}^{N} (y - \bar{x}\bar{a}) x_1 < \lambda(2-y)$$

Now we want to also consider the $-e_1$ direction:

$$\lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \left[ (y-1)(a_1 - \varepsilon)^2/2 + (2-y)|a_1 - \varepsilon| - (y-1)a_1^2/2 - (2-y)|a_1| \right]$$

$$\lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \left[ \frac{(y-1)}{2} \left[ a_1^2 - 2a_1\varepsilon + \varepsilon^2 - a_1^2 \right] + (2-y) \left[ |a_1 - \varepsilon| - |a_1| \right] \right]$$

$$\lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \left[ \frac{(y-1)}{2} \left[ -2a_1\varepsilon + \varepsilon^2 \right] + (2-y) \left( |a_1 - \varepsilon| - |a_1| \right] \right]$$

$$\lim_{\varepsilon \to 0} \frac{(y-1)}{2} \left( -2a_1 + \varepsilon \right] + (2-y) \left[ |a_1/\varepsilon - 1| - |a_1/\varepsilon| \right]$$

when $a_1 = 0$

$$= \lim_{\varepsilon \to 0} \frac{(y-1)}{2} (\varepsilon) + (2-y)|-1| = (2-y)$$

so

$$D_{e_{-1}} = \langle \nabla R, -e_1 \rangle + \lambda(2-y) \;:>0$$

$$= \frac{2}{N} \sum_{i=1}^{N} (y - \bar{x}\bar{a}) x_1 + \lambda(2-y)$$ This quantity is greater than 0 when

$$\frac{2}{N} \sum_{i=1}^{N} (y - \bar{x}\bar{a}) x_1 + \lambda(2-y) > 0$$

$$\frac{2}{N} \sum_{i=1}^{N} (y - \bar{x}\bar{a}) x_1 > -\lambda(2-y)$$

So as long as

$$-\lambda(2-\eta) < \frac{2}{N}\sum_{i=1}^{N}(y-\bar{x}\bar{a})\,x_1 < \lambda(2-\eta)\ ,$$

the minimum can occur when $a_1 = 0$. This can happen with $P > 0$ for any arbitrary $\lambda$, so there would be some $a_j$ s.t. $a_j^* = 0$.