Problem Set 5

Claire Goeckner-Wald

October 31, 2016

Linear Regression Error

1. [c] 100

$$
\begin{aligned}
\sigma &= 0.1 \\
d &= 8 \\
E_{in} &> 0.008 \\
\mathbb{E}_{\mathcal{D}}[E_{\text{in}}(\mathbf{w}_{\text{lin}})] &> \sigma^2(1 - \frac{d+1}{N}) \\
\frac{\mathbb{E}_{\mathcal{D}}[E_{\text{in}}(\mathbf{w}_{\text{lin}})]}{\sigma^2} &> 1 - \frac{d+1}{N} \\
\frac{d+1}{N} &> 1 - \frac{\mathbb{E}_{\mathcal{D}}[E_{\text{in}}(\mathbf{w}_{\text{lin}})]}{\sigma^2} \\
N &> \frac{d+1}{1 - \frac{\mathbb{E}_{\mathcal{D}}[E_{\text{in}}(\mathbf{w}_{\text{lin}})]}{\sigma^2}} \\
N &> \frac{8+1}{1 - \frac{0.008}{0.1^2}} \\
N &> 45
\end{aligned}
$$

Nonlinear Transforms

2. [D] $\widetilde{w}_1 < 0, \widetilde{w}_2 > 0$

$$
\begin{aligned}
\Phi(1, x_1, x_2) &= (1, x_1^2, x_2^2) \\
\text{sign}(\widetilde{\mathbf{w}}^T \cdot \widetilde{x}) &= \text{sign}(\widetilde{w}_0 \cdot 1 + \widetilde{w}_1 \cdot \widetilde{x}_1 + \widetilde{w}_2 \cdot \widetilde{x}_2)
\end{aligned}
$$

As $\widetilde{x}_1$ increases, the function should be more likely to return a negative number. Thus, $\widetilde{w}_1$ should be negative as well. As $\widetilde{x}_2$ increases, the function should be more likely to return a positive number. Thus, $\widetilde{w}_2$ should be positive as well.

3. [D] 20

The $d_{vc} \leq d + 1$. Here, $d = 15$, so $d_{vc} \leq 16$. Since none of the additional values are linear combinations of x_1 and x_2, then they are each new dimensions.

Gradient Descent

4. [D] $\widetilde{w}_1 < 0; \widetilde{w}_2 > 0$

$$\begin{aligned} E(u,v) &= (ue^v - 2ve^{-u})^2 \\ \frac{\partial E}{\partial u} &= \frac{\partial (ue^v - 2ve^{-u})^2}{\partial u} \\ &= 2(ue^v - 2ve^{-u}) \cdot (e^v + 2ve^{-u}) \end{aligned}$$

5. [D] 10

See attached code.

6. [E] (0.045, 0.024)

See attached code.

7. [A] 10^{-1}

See attached code.

Logistic Regression

8. [E] 0.125

See attached code.

9. [A] 350

See attached code.

PLA as SGD

10. [B] $-y_n \mathbf{w}^T \mathbf{x}_n$

PLA and SGD are similar because they use one point at a time to decrease the error. Although the minimization of error may cause small increases in error on other points, the average error reduces over time.