Machine Learning A1

Kody Manastyrski

September 2021

1 Question 1

1.1

$$J(\Theta) = \frac{h_\Theta(\hat{x}) - y}{2}^2$$

$$U(\Theta) = \Theta - \vec{\alpha} \sum_i^n \left[h_\Theta(x^{(i)}) - y^{(i)} \right] x^{(i)}$$

where $\vec{\alpha}$ is defined as the vector $\begin{bmatrix} \alpha \\ \alpha \\ \alpha \\ \alpha \end{bmatrix}$,

1.2

As iterations increase, the accuracy of the curve increases. Alpha values between 0.01 and 0.5 have positive impacts on the learning, allowing marginally closer to optimal results. To use an analogy, the iterations act like a course adjustment knob on a microscope, while the alpha value acts like a fine adjustment knob.

1.3

When the value increases past a certain point accuracy begins to fall off in the resulting curve. For the larger datasets, it seems that a dimensionality of 5 has the best results (from what we are allowed to experiment with), while smaller datasets only need dimensionality of 3 to have a good result, and anymore creates inaccuracy.

1.4

The previous results had good course accuracy, but clearly with the cosine element included, the optimal result is much closer to what the (hitherto unknown to experimenters) actual optimal best fit.

1.5

As mentioned above, when the dimensionality exceeds a certain point it accrues inaacuracy in the curve. In the case of smaller data sets, the inaccuracy happens much sooner, and the cosine term only serves to extend the number of dimensions available for fitting by 1.

2

2.1

$$\begin{aligned}
P(y=1) &= P(y=1|t=1,x)P(t=1) + P(y=1|t=0,x)P(t=0) \\
&= \alpha P(t=1) + 0 * P(t=0) \\
&= \alpha P(t=1) \\
P(t=1|y=1,x) &= \frac{P(y=1|t=1,x)P(t=1)}{P(y=1)} \\
&= \frac{\alpha P(t=1)}{P(y=1)} \\
&= \frac{P(y=1)}{P(y=1)} \\
&= 1
\end{aligned}$$

2.2

Given that X and T are independant and identically distributed, then we have

$$\begin{aligned}
P(t=1|x) &= P(t=1) \text{ and thusly} \\
\frac{1}{\alpha}P(y=1|x) &= P(t=1|x) \text{ becomes} \\
\frac{1}{\alpha}P(y=1|x) &= P(t=1) \\
P(y=1|x) &= \alpha P(t=1) \\
&= P(y=1)(\text{From above})
\end{aligned}$$

2.3

Proof of $P(h(x)|y=1)=\alpha$

$$
\begin{aligned}
&P(h(x)|y=1)\\
&=P((y=1|x)|y=1)\\
&=\frac{P(y=1|y=1|x)P(Y=1|x)}{P(y=1)}\\
&=\frac{P(y=1\cap y=1|x)P(y=1|x)}{P(y=1)P(y=1|x)}\\
&=\frac{P(y=1|x)P(y=1|x)}{P(y=1)P(y=1|x)}\\
&=\frac{1}{P(y=1)}\\
&=\frac{1}{\alpha P(t=1)}\\
&=\frac{P(t=1|y=1,x)}{\alpha P(t=1)}\\
&=\frac{P(y=1|t=1,x)P(t=1|x)}{\alpha P(y=1)P(t=1)}\\
&=\frac{\alpha P(t=1|x)}{\alpha P(y=1)P(t=1)}\\
&=\frac{\alpha P(y=1|x)}{P(y=1)P(t=1)}\\
&=\frac{\alpha P(x|y=1)}{P(x)P(t=1)}\\
&=\frac{\alpha P(x|y=1)}{P(x|t=1)}\\
&=\frac{\alpha P(y=1|x)P(t=1|x)}{P(t=|x)P(y=1|x)}\\
&=\alpha
\end{aligned}
$$

Proof of $P(h(x)|y=0)=0$

$$\begin{aligned}
&P(h(x)|y=0)\\
&=P((y=1|x)|y=0)\\
&=\frac{P(y=1|x\cap y=0}{P(y=0)}\\
&=\frac{P(y=0\cap y=1|x}{P(y=0)}\\
&=\frac{P(0|x)}{P(y=0)}
\end{aligned}$$