DSC 140A - Homework 08

Due: Wednesday, March 8

Note! Because the midterm is on Thursday, we would like to release the solutions to this homework immediately after the due date on Wednesday at midnight. Since this is before the late due date, we **cannot accept slip days for this homework**.

Problem 1.

You've been hired by a generic online retailer named after a rainforest named after a river. Your job is to build a model to predict whether or not a particular item will sell. You are provided with a dataset of outcomes for a collection of products:

Brand	Price Range	Condition	Sold
A	High	Used	No
A	High	New	Yes
B	Low	New	Yes
C	Medium	New	Yes
B	Low	Used	No
A	High	New	No
C	High	Used	Yes
A	Medium	Used	Yes
B	Medium	Used	No
C	Low	New	No
B	Low	Used	Yes

Using a Naïve Bayes classifier and the data above, predict if a product with Brand = B, Price Range = Medium, Condition = Used will sell or not. Show your calculations.

Solution:

We start by calculating the class conditional probabilities:

$$P(\text{Brand} = \text{B} \,|\, \text{Sold} = \text{Yes}) = \frac{2}{6}$$

$$P(\text{Brand} = \text{B} \,|\, \text{Sold} = \text{No}) = \frac{2}{5}$$

$$P(\text{Price Range} = \text{Medium} \,|\, \text{Sold} = \text{Yes}) = \frac{2}{6}$$

$$P(\text{Price Range} = \text{Medium} \,|\, \text{Sold} = \text{No}) = \frac{1}{5}$$

$$P(\text{Condition} = \text{Used} \,|\, \text{Sold} = \text{Yes}) = \frac{3}{6}$$

$$P(\text{Condition} = \text{Used} \,|\, \text{Sold} = \text{No}) = \frac{3}{5}$$

The prior probabilities are $P(\text{Sold} = \text{Yes}) = 6/11$ and $P(\text{Sold} = \text{No}) = 5/11$. Therefore:

$$
\begin{aligned}
&P(\text{Sold} = \text{Yes} \,|\, \text{Brand} = \text{B}, \text{Price Range} = \text{Medium}, \text{Condition} = \text{Used}) \\
&\qquad \propto \frac{2}{6} \cdot \frac{2}{6} \cdot \frac{3}{6} \cdot \frac{6}{11} \\
&= \frac{72}{2376} \approx 0.03
\end{aligned}
$$

$$
\begin{aligned}
&P(\text{Sold} = \text{No} \,|\, \text{Brand} = \text{B}, \text{Price Range} = \text{Medium}, \text{Condition} = \text{Used}) \\
&\qquad \propto \frac{2}{5} \cdot \frac{1}{5} \cdot \frac{3}{5} \cdot \frac{5}{11} \\
&= \frac{30}{1375} \approx 0.021
\end{aligned}
$$

Because the former is larger, we predict that the product will be **sold**.

Problem 2.

In lecture, we discussed the regression model where the target, Y, has the distribution

$$Y \sim \mathcal{N}(\text{Aug}(\vec{x}) \cdot \vec{w}, \sigma^2).$$

We also wrote down the log-likelihood of the parameters $\vec{w}$ and σ with respect to a data set of n points, $\{(\vec{x}^{(i)}, y_i)\}$:

$$\tilde{L}(\vec{w}, \sigma) = -\frac{1}{2\sigma^2} \sum_{i=1}^{n} \left(\mathsf{Aug}(\vec{x}^{(i)}) \cdot \vec{w} - y_i \right)^2 + \frac{n}{2} \ln \frac{1}{\sigma^2} - \frac{n}{2} \ln(2\pi)$$

Show that the maximum likelihood estimator for σ is:

$$\sigma_{\text{MLE}} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left[\mathsf{Aug}(\vec{x}^{(i)}) \cdot \vec{w}_{\text{MLE}} - y_i \right]^2},$$

where w_{MLE} is the maximum likelihood estimator for $\vec{w}$ (you do not need to re-derive it).

Solution: We start by taking a derivative of the log likelihood with respect to σ:

$$
\begin{aligned}
\frac{d}{d\sigma} \tilde{L}(\vec{w}, \sigma) &= \frac{d}{d\sigma} \left[-\frac{1}{2\sigma^2} \sum_{i=1}^{n} \left(\mathsf{Aug}(\vec{x}^{(i)}) \cdot \vec{w} - y_i \right)^2 + \frac{n}{2} \ln \frac{1}{\sigma^2} - \frac{n}{2} \ln(2\pi) \right] \\
&= \frac{d}{d\sigma} \left[-\frac{1}{2\sigma^2} \sum_{i=1}^{n} \left(\mathsf{Aug}(\vec{x}^{(i)}) \cdot \vec{w} - y_i \right)^2 \right] + \frac{d}{d\sigma} \frac{n}{2} \ln \frac{1}{\sigma^2} - \frac{d}{d\sigma} \frac{n}{2} \ln(2\pi) \\
&= \frac{2}{2\sigma^3} \sum_{i=1}^{n} \left(\mathsf{Aug}(\vec{x}^{(i)}) \cdot \vec{w} - y_i \right)^2 + \frac{n}{2} \sigma^2 (-2) \sigma^{-3} - 0 \\
&= \frac{2}{2\sigma^3} \sum_{i=1}^{n} \left(\mathsf{Aug}(\vec{x}^{(i)}) \cdot \vec{w} - y_i \right)^2 - \frac{n}{\sigma}
\end{aligned}
$$

Setting to zero and solving for σ:

$$
\begin{aligned}
&\frac{2}{2\sigma^3}\sum_{i=1}^{n}\left(\mathsf{Aug}(\vec{x}^{(i)})\cdot\vec{w}-y_i\right)^2-\frac{n}{\sigma}=0\\
\implies &\frac{1}{\sigma^2}\sum_{i=1}^{n}\left(\mathsf{Aug}(\vec{x}^{(i)})\cdot\vec{w}-y_i\right)^2-n=0\\
\implies &\frac{1}{n}\sum_{i=1}^{n}\left(\mathsf{Aug}(\vec{x}^{(i)})\cdot\vec{w}-y_i\right)^2=\sigma^2\\
\implies &\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\mathsf{Aug}(\vec{x}^{(i)})\cdot\vec{w}-y_i\right)^2}=\sigma
\end{aligned}
$$

We can substitute the optimal choice of $\vec{w}=\vec{w}_{\text{MLE}}$ in order to arrive at the final answer.