Assignment 1: Machine learning basics

Jan Scheffczyk - 3242317
Sarah Khan - 3279206
Mahmoud Hashem - 3201329
Mohamed Saleh - 3201337

October 23, 2019

1 Theoretical exercises

1.1 Compute bias of estimators

$$\begin{aligned}
\text{Bias}\left(\hat{\sigma}_{\text{m}}^2\right) =& E\left[\frac{1}{m}\sum_{i=1}^{m}\left(x^{(i)} - \hat{\mu_m}\right)^2\right] - \sigma^2 \\
=& \frac{1}{m}\left(m-1\right)\sigma^2 - \sigma^2 \\
=& \left(\frac{1}{m}\left(m-1\right) - 1\right)\sigma^2 \\
=& -\frac{1}{m}\sigma^2
\end{aligned}$$

The estimates are of the variance $\hat{\sigma}_m^2$ is biased.

$$\begin{aligned}
\text{Bias}\left(\tilde{\sigma}_{\text{m}}^2\right) =& E\left[\frac{1}{m-1}\sum_{i=1}^{m}\left(x^{(i)} - \hat{\mu}_m\right)^2\right] \\
=& \frac{1}{m-1}(m-1)\sigma^2 - \sigma^2 \\
=& 0
\end{aligned}$$

The estimates are of the variance $\tilde{\sigma}_m^2$ is unbiased.

1.2 Bias Variance Trade-off

Show that $\mathrm{MSE} = \mathrm{Bias}^2 + \mathrm{Var}$

$$
\begin{aligned}
\mathrm{Bias}^2\left(\hat{\theta}_{\mathrm{m}}\right) =& \left(E\left[\hat{\theta}_m\right] - \theta\right)^2 \\
=& E^2\left[\hat{\theta}_m\right] - 2E\left[\hat{\theta}\right]\theta + \theta^2 \\
\mathrm{Var}\left(\hat{\theta}_{\mathrm{m}}\right) =& E\left[\hat{\theta}_m^2\right] - E^2\left[\hat{\theta}_m\right] \\
\mathrm{MSE}\left(\hat{\theta}_{\mathrm{m}}\right) =& E\left[\left(\hat{\theta}_m - \theta\right)^2\right] \\
=& E\left[\hat{\theta}_m^2\right] - 2E\left[\hat{\theta}_m\right]\theta + \theta^2 \\
=& E\left[\hat{\theta}_m^2\right] - 2E\left[\hat{\theta}_m\right]\theta + \theta^2 + \left(E\left[\hat{\theta}_m\right] - E^2\left[\hat{\theta}_m\right]\right) \\
=& \underbrace{E^2\left[\hat{\theta}_m\right] - 2E\left[\hat{\theta}_m\right]\theta + \theta^2}_{\mathrm{Bias}^2} + \underbrace{E\left[\hat{\theta}_m^2\right] - E^2\left[\hat{\theta}_m\right]}_{\mathrm{Var}}
\end{aligned}
$$

1.3 Maximum a posteriori

$$
P\left(\theta|x,y\right) = \frac{P\left(y|x,\theta\right)P\left(\theta|x\right)}{P\left(x\right)} \propto P\left(y|x,\theta\right)P\left(\theta|x\right)
$$

Since θ this independent of x this is equivalent to $P\left(y|x,\theta\right)P\left(\theta\right)$. Thus we get the maximum a posteriori approximation:

$$
\theta_{\mathrm{MAP}} = \underset{\theta}{\operatorname{argmax}}\left[\sum_{i=1}^{m}\log\left(P\left(y^{(i)}|x^{(i)},\theta\right) + \log\left(P\left(\theta\right)\right)\right)\right]
$$