NEW SHINY APP FORMULAE
April 25, 2024

BERTRAND NORTIER, DANIEL WILLIAMSON

1. Current method

We assume that we have a model with input $\mathbf{x}$ and scalar output $Y(\mathbf{x})$. We assume that $Y(\mathbf{x})$ is modelled as a (Deep Gaussian Process). We also know that $Y(\mathbf{x}) \geq 0; \forall \mathbf{x}$. Given a target $\mathcal{T}$, we want to estimate: $P\left(Y\left(\mathbf{x}\right) \geq \mathcal{T}\right)$. If we make the assumption that $\mathcal{T} \geq E[Y(\mathbf{x})]$, we can use Cantelli's inequality to derive an upper bound on this probability.

$$P\left(\left(Y\left(\mathbf{x}\right) - E\left[Y\left(\mathbf{x}\right)\right]\right) \geq \left(\mathcal{T} - E\left[Y\left(\mathbf{x}\right)\right]\right) | \mathcal{T} \geq E\left[Y\left(\mathbf{x}\right)\right]\right) \leq \frac{Var\left[Y\left(\mathbf{x}\right)\right]}{Var\left[Y\left(\mathbf{x}\right)\right] + \left(\mathcal{T} - E\left[Y\left(\mathbf{x}\right)\right]\right)^2}$$

writing the Implausibility as:

$$I\left(\mathbf{x}\right) = \frac{\mathcal{T} - E\left[Y\left(\mathbf{x}\right)\right]}{\sqrt{Var\left[Y\left(\mathbf{x}\right)\right]}}$$

and replacing, we obtain:

$$\boxed{P\left(Y\left(\mathbf{x}\right) \geq \mathcal{T} | \mathcal{T} \geq E\left[Y\left(\mathbf{x}\right)\right]\right) = P\left(\frac{\left(Y\left(\mathbf{x}\right) - E\left[Y\left(\mathbf{x}\right)\right]\right)}{\sqrt{Var\left[Y\left(\mathbf{x}\right)\right]}} \geq I\left(\mathbf{x}\right) | I\left(\mathbf{x}\right) \geq 0\right) \leq \frac{1}{1 + I\left(\mathbf{x}\right)^2}} \quad (1)$$

Similary, for cases where $\mathcal{T} \leq E[Y(\mathbf{x})]$, we can apply the negative side Cantelli's inequality:

$$\begin{aligned} P\left(Y\left(\mathbf{x}\right) \leq \mathcal{T} | \mathcal{T} \leq E\left[Y\left(\mathbf{x}\right)\right]\right) &= P\left(\left(Y\left(\mathbf{x}\right) - E\left[Y\left(\mathbf{x}\right)\right]\right) \leq \left(\mathcal{T} - E\left[Y\left(\mathbf{x}\right)\right]\right) | \mathcal{T} \leq E\left[Y\left(\mathbf{x}\right)\right]\right) \\ &\leq \frac{Var\left[Y\left(\mathbf{x}\right)\right]}{Var\left[Y\left(\mathbf{x}\right)\right] + \left(\mathcal{T} - E\left[Y\left(\mathbf{x}\right)\right]\right)^2} \end{aligned}$$

replacing by $I(\mathbf{x})$, we obtain:

$$\boxed{P\left(Y\left(\mathbf{x}\right) \leq \mathcal{T} | \mathcal{T} \leq E\left[Y\left(\mathbf{x}\right)\right]\right) = P\left(\frac{\left(Y\left(\mathbf{x}\right) - E\left[Y\left(\mathbf{x}\right)\right]\right)}{\sqrt{Var\left[Y\left(\mathbf{x}\right)\right]}} \leq -\left|I\left(\mathbf{x}\right)\right| | I\left(\mathbf{x}\right) \leq 0\right) \leq \frac{1}{1 + I\left(\mathbf{x}\right)^2}} \quad (2)$$

We then have:

$$\boxed{P\left(Y\left(\mathbf{x}\right) > \mathcal{T} | \mathcal{T} \leq E\left[Y\left(\mathbf{x}\right)\right]\right) = 1 - P\left(Y\left(\mathbf{x}\right) \leq \mathcal{T} | \mathcal{T} \leq E\left[Y\left(\mathbf{x}\right)\right]\right) \geq 1 - \frac{1}{1 + I\left(\mathbf{x}\right)^2} \geq \frac{I\left(\mathbf{x}\right)^2}{1 + I\left(\mathbf{x}\right)^2}} \quad (3)$$

The issue is that for the app, we must have the possibility to have both $\mathcal{T} \leq E[Y(\mathbf{x})]$ and $\mathcal{T} \geq E[Y(\mathbf{x})]$ and we see that first one bound obtained is an upper bound whilst the second bound is a lower bound. Second, the bounds are not continuous in 0 as we have: $\lim_{I(x)\to 0^-} \frac{I(\mathbf{x})^2}{1+I(\mathbf{x})^2} = 0$ and $\lim_{I(x)\to 0^+} \frac{1}{1+I(\mathbf{x})^2} = 1$ so we cannot use these 2 bounds in the app, otherwise there would be inconsistencies if the user crosses a threshold by moving the target above or below $E[Y(\mathbf{x}]$. We also note that if $\mathcal{T}$ is near $E[Y(\mathbf{x})]$ the 2 thresholds are simply implying that the probability is between 0 and 1 that are not sufficiently tight bounds to be used. We propose instead to make the stronger assumption that $Y(\mathbf{x})$ follows a Gaussian truncated at 0.

Date: April 25, 2024.

2. Truncated Gaussian

We first standardize $Y(\mathbf{x})$:

$$\overleftarrow{Y}(\mathbf{x}) = \frac{Y(\mathbf{x}) - E[Y(\mathbf{x})]}{\sqrt{Var[Y(\mathbf{x})]}}$$

We then note that:

$$P(Y(\mathbf{x}) \geq \mathcal{T}) = P\left(\overleftarrow{Y}(\mathbf{x}) \geq \mathcal{I}(\mathbf{x})\right)$$

If we assume that $Y(\mathbf{x})$ is Normally distributed, we could assume that $\overleftarrow{Y}(\mathbf{x})$ is a standard Gaussian (mean=0, var=1). However, we also know that we must have $Y(\mathbf{x}) \geq 0; \forall \mathbf{x}$. We therefore assume that

$$\overleftarrow{Y}(\mathbf{x})|\mathbf{x} \sim \mathcal{N}_{[a,\infty)}(0,1)$$

To determine a, we simply use the positivity of $Y(\mathbf{x})$:

$$Y(\mathbf{x}) \geq 0 \Leftrightarrow \overleftarrow{Y}(\mathbf{x}) \geq \frac{-E[Y(\mathbf{x})]}{\sqrt{Var[Y(\mathbf{x})]}}$$

Hence we assume that:

$$\overleftarrow{Y}(\mathbf{x})|\mathbf{x} \sim \mathcal{N}_{\left[\frac{-E[Y(\mathbf{x})]}{\sqrt{Var[Y(\mathbf{x})]}},\infty\right)}(0,1)$$

$$P(Y(\mathbf{x}) \geq \mathcal{T}) = 1 - P\left(\overline{Y}(\mathbf{x}) \leq I(\mathbf{x})\right) = 1 - \int_{\frac{-E[Y(\mathbf{x})]}{\sqrt{Var[Y(\mathbf{x})]}}}^{I(\mathbf{x})} \phi(x)\,dx = 1 - \Phi_{\left[\frac{-E[Y(\mathbf{x})]}{\sqrt{Var[Y(\mathbf{x})]}},\infty\right)}(I(\mathbf{x})) \quad (4)$$

A tolerance (for model error,...) ‘ be added to the variance at the denominator for the calculation of the Implausibility. The tolerance can be added to denominator of the lower bound $a = -\frac{E[Y(\mathbf{x})]}{\sqrt{(Var(Y(\mathbf{x})))}}$. This avoids infinite values for $|I(\mathbf{x})|$ and $-a$ if the variance of $Y(\mathbf{x})$ is 0.

2.1. Max Target Initialization.

Let's now assume that we have a known confidence level α that we want to achieve:

$$P(Y(\mathbf{x}) \geq \mathcal{T}) \geq \alpha$$

and we want to determine the maximum possible value allowed for $\mathcal{T}$ such that if the parcels are all planted, then the Target is met with a probability at least α:

$$P(Y(\mathbf{x}_{all}) \geq \mathcal{T}) \geq \alpha \Leftrightarrow 1 - \Phi_{\left[\frac{-E[Y(\mathbf{x}_{all})]}{\sqrt{Var[Y(\mathbf{x}_{all})]+tol}},\infty\right)}\left(\frac{\mathcal{T}_{max} - E[Y(\mathbf{x}_{all})]}{\sqrt{Var[Y(\mathbf{x}_{all})]+tol}}\right) \geq \alpha$$

$$\boxed{\left(\sqrt{Var[Y(\mathbf{x}_{all})]+tol}\right) \times \left(\Phi^{-1}_{\left[\frac{-E[Y(\mathbf{x}_{all})]}{\sqrt{Var[Y(\mathbf{x}_{all})]+tol}},\infty\right)}(1-\alpha)\right) + E[Y(\mathbf{x}_{all})] \geq \mathcal{T}_{max}} \quad (5)$$

2.2. Formulae for $\leq$.

We have:

$$P(Y(\mathbf{x}) \leq \mathcal{T}) = P\left(\overleftarrow{Y}(\mathbf{x}) \leq \mathcal{I}(\mathbf{x})\right) = \int_{-\frac{E[Y(\mathbf{x})]}{\sqrt{var[Y(\mathbf{x})]}}}^{I(\mathbf{x})} \phi(u)\,du = \Phi_{\left[-\frac{E[Y(\mathbf{x})]}{\sqrt{var[Y(\mathbf{x})]}},\infty\right]}(I(\mathbf{x}))$$

$$P(Y(\mathbf{x}_{min}) \leq \mathcal{T}) \geq \alpha \Leftrightarrow P\left(\overleftarrow{Y}(\mathbf{x}_{min}) \leq \mathcal{I}(\mathbf{x}_{min})\right) \geq \alpha \Leftrightarrow \Phi_{\left[-\frac{E[Y(\mathbf{x}_{min})]}{\sqrt{var[Y(\mathbf{x}_{min})]}},\infty\right]}\left(\frac{\mathcal{T}_{min} - E[Y(\mathbf{x}_{min})]}{\sqrt{var[Y(\mathbf{x}_{min})]}}\right) \geq \alpha$$

$$\mathcal{T}_{min} \geq E[Y(\mathbf{x}_{min})] + \sqrt{var[Y(\mathbf{x}_{min})]}\left(\Phi^{-1}_{\left[-\frac{E[Y(\mathbf{x}_{min})]}{\sqrt{var[Y(\mathbf{x}_{min})]}},\infty\right]}(\alpha)\right)$$

In the case x is the planting area, $\mathbf{x}_{min}$ consists in no planting and the area is zero. In this case $E[Y(x_{min})] = 0$. In practice, we add a tolerance to the variance. Hence even if $varl[Y(x_{min})] = 0$, we still have a tolerance in the square root of the second argument. Hence:

$$\mathcal{T}_{min}^{area} \geq \sqrt{tol_{area}} \left(\Phi_{[0,\infty]}^{-1} \left(\alpha \right) \right)$$

We also need to have:

$$P\left(Y\left(\mathbf{x}_{max}\right) \leq \mathcal{T}_{max}\right) \geq \alpha \Leftrightarrow P\left(\overleftarrow{Y}\left(\mathbf{x}_{max}\right) \leq \mathcal{I}\left(\mathbf{x}_{max}\right)\right) \geq \alpha \Leftrightarrow \Phi_{\left[-\frac{E[Y(\mathbf{x}_{max})]}{\sqrt{var[Y(\mathbf{x}_{max})]}},\infty\right]}\left(\frac{\mathcal{T}_{max} - E\left[Y\left(\mathbf{x}_{max}\right)\right]}{\sqrt{var\left[Y\left(\mathbf{x}_{max}\right)\right]}}\right) \geq \alpha$$

$$\mathcal{T}_{max} \geq \sqrt{var\left[Y\left(\mathbf{x}_{max}\right)\right]}\left(\Phi^{-1}_{\left[-\frac{E[Y(\mathbf{x}_{max})]}{\sqrt{var[Y(\mathbf{x}_{max})]}},\infty\right]}\left(\alpha\right)\right) + E\left[Y\left(\mathbf{x}_{max}\right)\right]$$