

Сразу обозначим следующую закономерность. Из рисунка 2.2 сверху можно сделать вывод:

$$\begin{aligned} FPR(PD') &= 1 - \hat{F}_P(PD'), \\ TPR(PD') &= 1 - \hat{F}_N(PD'). \end{aligned}$$

Используя эту закономерность вместе с (2.3), получим:

$$KS = \max_{PD'} |FPR(PD') - TPR(PD')|.$$

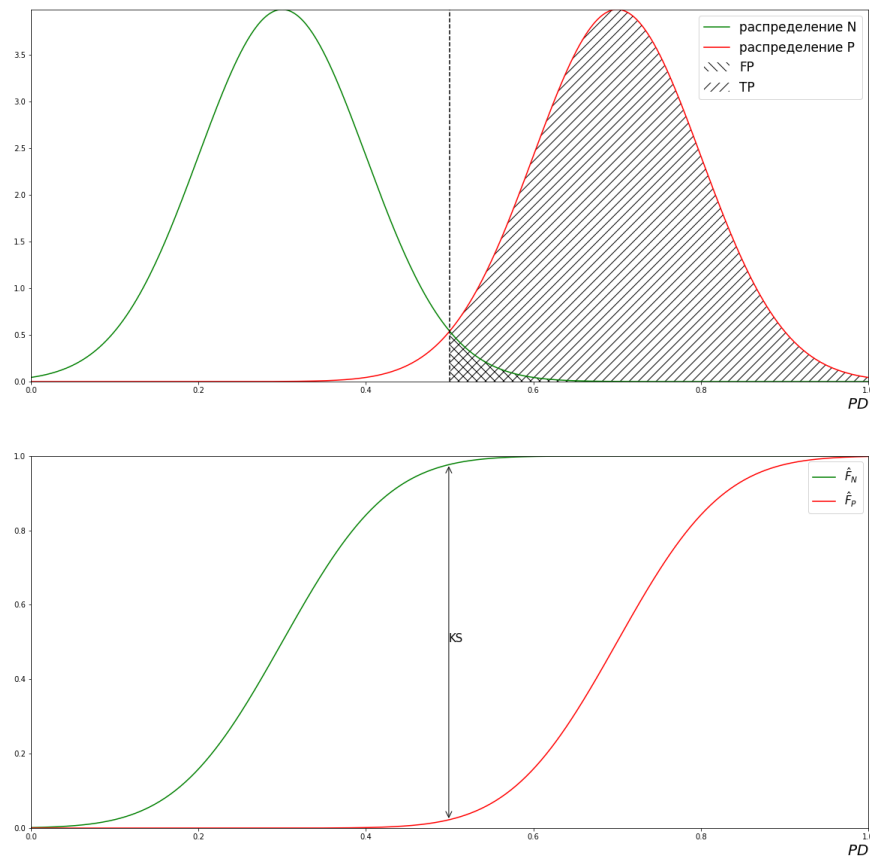


Рисунок 2.2 – Смысл KS статистики

Примечание – Источник: собственная разработка.

Итак, KS это еще и максимальная разница долей истинно положительных и ложно положительных предсказаний модели. Отсюда очень важный для нас в дальнейшем вывод – точка PD^* , соответствующая KS статистике, тот вариант точки отсечения который позволяет максимально нарастить долю правильных предсказаний исследуемого признака при минимально допустимой доле ошибочных предсказаний. Формально такую точку можно записать:

$$\begin{aligned} PD^* &= \operatorname{argmax}_{PD'} |\hat{F}_P(PD') - \hat{F}_N(PD')| = \\ &= \operatorname{argmax}_{PD'} |FPR(PD') - TPR(PD')|. \end{aligned} \tag{2.4}$$