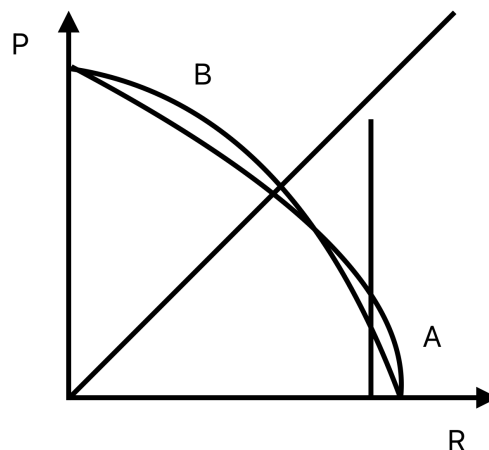


1. 若学习器 A 的 F1 值比学习器 B 的 F1 值高，试分析 A 的 EBP 值是否也比 B 高

$$F_1 = \frac{2}{\frac{1}{R} + \frac{1}{P}} = \frac{2PR}{P + R}$$

F1_score 比较高说明学习的效果比较稳定，精确率和召回率都处于一个比较好的状态，EBP 值表示在 PR 曲线中 $P = R$ 的情况对应值，如图所示 $EBP_A < EBP_B$ 。而分类阈值会影响 F1_score 的大小，在曲线的右侧，有可能出现选取的阈值结果 $F1_score_A > F1_score_B$ 的情况，因为 P 和 R 决定着 F1 的取值，相同的 R 的情况下，P 越大 F1 越大，如图所示，在曲线 A 的精确率 P 要比 B 的 P 大，所以对应的 F1 大，所以当学习器 A 的 F1 值比学习器 B 的 F1 值高时 A 的 EBP 值不一定比 B 高



2. 试述真正利率、假正利率与查准率、查全率之间的联系

真正利率：公式计算为 $TPR = TP / (TP + FN)$ 表示为在所有的正样本中被预测正确的概率，即被预测为正样本。

假正利率：公式计算为 $FPR = FP / (FP + TN)$ 表示在所有实际为阴性的样本中，被错误地判断为阳性之比率。

查准率：又被称为精确率，公式计算为 $precision = TP / (TP + FP)$ ，即表示在所有被预测为正确的样本中，真正被预测对的概率。

查全率：又被称为召回率，公式计算为 $recall = TP / (TP + FN)$ ，和真正利率表示的含义相同，表示在所有的正样本中被预测正确的概率。