

CLASSIFICADORES BAYESIANOS

Marcos Lopes

Departamento de Linguística

Classificadores

Classificadores bayesianos

Exemplo de aplicação

Avaliação

PROBLEMA GERAL DAS CLASSIFICAÇÕES

- As tarefas de classificação são aquelas em que dados (ou observações) são julgados como pertencentes a uma determinada classe c de um conjunto finito de classes \mathcal{C} .

PROBLEMA GERAL DAS CLASSIFICAÇÕES

- As tarefas de classificação são aquelas em que dados (ou observações) são julgados como pertencentes a uma determinada classe c de um conjunto finito de classes \mathcal{C} .
- Nos problemas ligados ao PLN, as classes são discretas (sem encavalamento) e complementares (suas respectivas probabilidades devem somar 1).

PROBLEMA GERAL DAS CLASSIFICAÇÕES

- As tarefas de classificação são aquelas em que dados (ou observações) são julgados como pertencentes a uma determinada classe c de um conjunto finito de classes \mathcal{C} .
- Nos problemas ligados ao PLN, as classes são discretas (sem encavalamento) e complementares (suas respectivas probabilidades devem somar 1).
- Há duas formas básicas de classificação:

PROBLEMA GERAL DAS CLASSIFICAÇÕES

- As tarefas de classificação são aquelas em que dados (ou observações) são julgados como pertencentes a uma determinada classe c de um conjunto finito de classes \mathcal{C} .
- Nos problemas ligados ao PLN, as classes são discretas (sem encavalamento) e complementares (suas respectivas probabilidades devem somar 1).
- Há duas formas básicas de classificação:
 - Classificação baseada em regras

PROBLEMA GERAL DAS CLASSIFICAÇÕES

- As tarefas de classificação são aquelas em que dados (ou observações) são julgados como pertencentes a uma determinada classe c de um conjunto finito de classes \mathcal{C} .
- Nos problemas ligados ao PLN, as classes são discretas (sem encavalamento) e complementares (suas respectivas probabilidades devem somar 1).
- Há duas formas básicas de classificação:
 - Classificação baseada em regras
 - Classificação baseada em aprendizado de máquina (supervisionado ou não)

CLASSIFICADORES PROBABILÍSTICOS

- Entre os algoritmos baseados no aprendizado de máquina supervisionado, destacam-se os classificadores probabilísticos.

CLASSIFICADORES PROBABILÍSTICOS

- Entre os algoritmos baseados no aprendizado de máquina supervisionado, destacam-se os classificadores probabilísticos.
- Sua característica central é que, além de classificar os dados em classes, eles são capazes de informar a probabilidade de inclusão dos dados nas classes, o que é muito útil para que se tenha uma ideia do grau de certeza do julgamento.

CLASSIFICADORES PROBABILÍSTICOS

- Entre os algoritmos baseados no aprendizado de máquina supervisionado, destacam-se os classificadores probabilísticos.
- Sua característica central é que, além de classificar os dados em classes, eles são capazes de informar a probabilidade de inclusão dos dados nas classes, o que é muito útil para que se tenha uma ideia do grau de certeza do julgamento.
 - Essa é uma vantagem importante por relação aos algoritmos baseados em regras.

CLASSIFICADORES PROBABILÍSTICOS

- Entre os algoritmos baseados no aprendizado de máquina supervisionado, destacam-se os classificadores probabilísticos.
- Sua característica central é que, além de classificar os dados em classes, eles são capazes de informar a probabilidade de inclusão dos dados nas classes, o que é muito útil para que se tenha uma ideia do grau de certeza do julgamento.
 - Essa é uma vantagem importante por relação aos algoritmos baseados em regras.
- Em linhas gerais, o procedimento de classificação é o seguinte:

CLASSIFICADORES PROBABILÍSTICOS

- Entre os algoritmos baseados no aprendizado de máquina supervisionado, destacam-se os classificadores probabilísticos.
- Sua característica central é que, além de classificar os dados em classes, eles são capazes de informar a probabilidade de inclusão dos dados nas classes, o que é muito útil para que se tenha uma ideia do grau de certeza do julgamento.
 - Essa é uma vantagem importante por relação aos algoritmos baseados em regras.
- Em linhas gerais, o procedimento de classificação é o seguinte:
 1. Para cada uma das possíveis classes c pertencentes a um conjunto pré-determinado \mathcal{C} , calcula-se a probabilidade de um documento d ser classificado em c : $P(c|d)$.

CLASSIFICADORES PROBABILÍSTICOS

- Entre os algoritmos baseados no aprendizado de máquina supervisionado, destacam-se os classificadores probabilísticos.
- Sua característica central é que, além de classificar os dados em classes, eles são capazes de informar a probabilidade de inclusão dos dados nas classes, o que é muito útil para que se tenha uma ideia do grau de certeza do julgamento.
 - Essa é uma vantagem importante por relação aos algoritmos baseados em regras.
- Em linhas gerais, o procedimento de classificação é o seguinte:
 1. Para cada uma das possíveis classes c pertencentes a um conjunto pré-determinado \mathcal{C} , calcula-se a probabilidade de um documento d ser classificado em c : $P(c|d)$.
 2. A partir desse cálculo, mapeia-se d na classe estimada (\hat{c}) como sendo a de máxima probabilidade: $\hat{c} = \underset{c \in \mathcal{C}}{\operatorname{argmax}} P(c|d)$

Classificadores

Classificadores bayesianos

Exemplo de aplicação

Avaliação

CLASSIFICADORES BAYESIANOS INGÊNUOS

- Um documento d pode ser concebido como um vetor de n atributos ou traços (*features*) f_1, \dots, f_n . Pense nos atributos como palavras ou outros elementos relevantes para a classificação.

CLASSIFICADORES BAYESIANOS INGÊNUOS

- Um documento d pode ser concebido como um vetor de n atributos ou traços (*features*) f_1, \dots, f_n . Pense nos atributos como palavras ou outros elementos relevantes para a classificação.
- Substituindo d na fórmula anterior, temos:

$$\hat{c} = \operatorname{argmax}_{c \in \mathcal{C}} P(c | f_1 \& f_2 \& \dots \& f_n) \quad (1)$$

CLASSIFICADORES BAYESIANOS INGÊNUOS

- Um documento d pode ser concebido como um vetor de n atributos ou traços (*features*) f_1, \dots, f_n . Pense nos atributos como palavras ou outros elementos relevantes para a classificação.
- Substituindo d na fórmula anterior, temos:

$$\hat{c} = \operatorname{argmax}_{c \in \mathcal{C}} P(c|f_1 \& f_2 \& \dots \& f_n) \quad (1)$$

- Aqui entra a Regra de Bayes, que permite calcular probabilidades condicionais como essa acima:

$$P(X|Y) = \frac{P(X) P(Y|X)}{P(Y)} \quad (2)$$

CLASSIFICADORES BAYESIANOS INGÊNUOS

- Um documento d pode ser concebido como um vetor de n atributos ou traços (*features*) f_1, \dots, f_n . Pense nos atributos como palavras ou outros elementos relevantes para a classificação.
- Substituindo d na fórmula anterior, temos:

$$\hat{c} = \underset{c \in \mathcal{C}}{\operatorname{argmax}} P(c | f_1 \& f_2 \& \dots \& f_n) \quad (1)$$

- Aqui entra a Regra de Bayes, que permite calcular probabilidades condicionais como essa acima:

$$P(X|Y) = \frac{P(X) P(Y|X)}{P(Y)} \quad (2)$$

- A Regra de Bayes permite calcular a probabilidade condicional de X , dado Y , a partir da probabilidade condicional de Y dado X e das probabilidades simples (não-condicionais) de X e Y .

CLASSIFICADORES BAYESIANOS INGÊNUOS (CONT.)

- Aplicando a Regra de Bayes, temos:

$$\hat{c} = \operatorname{argmax}_{c \in \mathcal{C}} \frac{P(c) P(f_1 \& f_2 \& \dots \& f_n | c)}{P(f_1 \& f_2 \& \dots \& f_n)} \quad (3)$$

CLASSIFICADORES BAYESIANOS INGÊNUOS (CONT.)

- Aplicando a Regra de Bayes, temos:

$$\hat{c} = \operatorname{argmax}_{c \in \mathcal{C}} \frac{P(c) P(f_1 \& f_2 \& \dots \& f_n | c)}{P(f_1 \& f_2 \& \dots \& f_n)} \quad (3)$$

- O denominador da fração acima será sempre o mesmo para todas as classes, pois ele representa o documento a classificar, não a classe. Assim, para nosso objetivo de classificação, podemos eliminá-lo:

$$\hat{c} = \operatorname{argmax}_{c \in \mathcal{C}} P(c) P(f_1 \& f_2 \& \dots \& f_n | c) \quad (4)$$

CLASSIFICADORES BAYESIANOS INGÊNUOS (CONT.)

- Aplicando a Regra de Bayes, temos:

$$\hat{c} = \operatorname{argmax}_{c \in \mathcal{C}} \frac{P(c) P(f_1 \& f_2 \& \dots \& f_n | c)}{P(f_1 \& f_2 \& \dots \& f_n)} \quad (3)$$

- O denominador da fração acima será sempre o mesmo para todas as classes, pois ele representa o documento a classificar, não a classe. Assim, para nosso objetivo de classificação, podemos eliminá-lo:

$$\hat{c} = \operatorname{argmax}_{c \in \mathcal{C}} P(c) P(f_1 \& f_2 \& \dots \& f_n | c) \quad (4)$$

- Por fim, considerando-se a **hipótese ingênua de Bayes** de que as probabilidades dos traços são independentes entre si, temos:

$$\hat{c} = \operatorname{argmax}_{c \in \mathcal{C}} P(c) \prod_{i=1}^n P(f_i | c) \quad (5)$$

Classificadores

Classificadores bayesianos

Exemplo de aplicação

Avaliação

EXEMPLO DE APLICAÇÃO: DETECÇÃO DE SPAM

- Uma das tarefas mais comuns em que os classificadores bayesianos são empregados é a detecção de spams.

EXEMPLO DE APLICAÇÃO: DETECÇÃO DE SPAM

- Uma das tarefas mais comuns em que os classificadores bayesianos são empregados é a detecção de spams.
- Nessa tarefa, os dados de entrada são mensagens (em geral de e-mail) que devem ser classificadas em duas classes:

$$\mathcal{C} = \{\text{spam}, \text{ham}\}$$

EXEMPLO DE APLICAÇÃO: DETECÇÃO DE SPAM

- Uma das tarefas mais comuns em que os classificadores bayesianos são empregados é a detecção de spams.
- Nessa tarefa, os dados de entrada são mensagens (em geral de e-mail) que devem ser classificadas em duas classes:

$$\mathcal{C} = \{\text{spam}, \text{ham}\}$$

- Os traços (ou “atributos”, *features*) usados na classificação virão, no nosso estudo, somente do conteúdo textual da mensagem.

EXEMPLO DE APLICAÇÃO: DETECÇÃO DE SPAM

- Uma das tarefas mais comuns em que os classificadores bayesianos são empregados é a detecção de spams.
- Nessa tarefa, os dados de entrada são mensagens (em geral de e-mail) que devem ser classificadas em duas classes:

$$\mathcal{C} = \{\text{spam}, \text{ham}\}$$

- Os traços (ou “atributos”, *features*) usados na classificação virão, no nosso estudo, somente do conteúdo textual da mensagem.
 - Na prática, os detectores de spam tendem a incorporar outros atributos da mensagem, como verificar se o remetente faz parte da lista de contatos do destinatário, se há muitos destinatários para a mensagem, se há imagens, links e outros objetos não textuais.

EXEMPLO DE APLICAÇÃO: DETECÇÃO DE SPAM

- Uma das tarefas mais comuns em que os classificadores bayesianos são empregados é a detecção de spams.
- Nessa tarefa, os dados de entrada são mensagens (em geral de e-mail) que devem ser classificadas em duas classes:

$$\mathcal{C} = \{\text{spam}, \text{ham}\}$$

- Os traços (ou “atributos”, *features*) usados na classificação virão, no nosso estudo, somente do conteúdo textual da mensagem.
 - Na prática, os detectores de spam tendem a incorporar outros atributos da mensagem, como verificar se o remetente faz parte da lista de contatos do destinatário, se há muitos destinatários para a mensagem, se há imagens, links e outros objetos não textuais.
- Como vamos trabalhar com conteúdo textual, cada mensagem deve ser pré-processada a fim de minimizar variabilidade e erro. No nosso exemplo, consideraremos já efetuadas as do pré-processamento.

EXEMPLO DE APLICAÇÃO (CONT.)

- Os dados para análise serão, portanto, vetores de palavras plenas f_1, \dots, f_n representando os atributos das mensagens.

EXEMPLO DE APLICAÇÃO (CONT.)

- Os dados para análise serão, portanto, vetores de palavras plenas f_1, \dots, f_n representando os atributos das mensagens.
- Os cálculos de probabilidades serão feitos através de estimativa por máxima verossimilhança (MLE), ou seja, calculando-se as frequências relativas dos traços e classes.

EXEMPLO DE APLICAÇÃO (CONT.)

- Os dados para análise serão, portanto, vetores de palavras plenas f_1, \dots, f_n representando os atributos das mensagens.
- Os cálculos de probabilidades serão feitos através de estimativa por máxima verossimilhança (MLE), ou seja, calculando-se as frequências relativas dos traços e classes.
- Seja N o número de documentos (mensagens) de nosso corpus.

EXEMPLO DE APLICAÇÃO (CONT.)

- Os dados para análise serão, portanto, vetores de palavras plenas f_1, \dots, f_n representando os atributos das mensagens.
- Os cálculos de probabilidades serão feitos através de estimativa por máxima verossimilhança (MLE), ou seja, calculando-se as frequências relativas dos traços e classes.
- Seja N o número de documentos (mensagens) de nosso corpus.
- A probabilidade isolada (“anterior”) da classe c é dada por:

$$P(c) = \frac{\text{contagem}(c)}{N} \quad (6)$$

EXEMPLO DE APLICAÇÃO (CONT.)

- Os dados para análise serão, portanto, vetores de palavras plenas f_1, \dots, f_n representando os atributos das mensagens.
- Os cálculos de probabilidades serão feitos através de estimativa por máxima verossimilhança (MLE), ou seja, calculando-se as frequências relativas dos traços e classes.
- Seja N o número de documentos (mensagens) de nosso corpus.
- A probabilidade isolada (“anterior”) da classe c é dada por:

$$P(c) = \frac{\text{contagem}(c)}{N} \quad (6)$$

- Já a probabilidade de um atributo qualquer pertencer a c é:

$$P(f_i|c) = \frac{\text{contagem}(f_i, c)}{\text{contagem}(c)} \quad (7)$$

EXEMPLO DE APLICAÇÃO: CALCULANDO

Considere que as seguintes contagens foram registradas em nosso corpus:

	spam	ham
N	6.000	4.000
viagra	3.600	40
milhões	2.400	1.000
reunião	90	120

EXEMPLO DE APLICAÇÃO: CALCULANDO (CONT.)

As probabilidades associadas às contagens são:

	spam	ham
N	0,6	0,4
viagra	0,6	0,01
milhões	0,4	0,25
reunião	0,015	0,03

EXEMPLO DE APLICAÇÃO: CALCULANDO (CONT.)

Imagine agora que se deseje classificar uma mensagem com as seguintes palavras:

viagra, reunião, chefia

EXEMPLO DE APLICAÇÃO: CALCULANDO (CONT.)

Imagine agora que se deseje classificar uma mensagem com as seguintes palavras:

viagra, reunião, chefia

A palavra “chefia” não aparece no vocabulário já modelizado. Por isso, ela será simplesmente descartada, como se a mensagem contivesse somente as palavras “viagra” e “reunião”.

De resto, a fórmula:

$$P(c) \prod_{i=1}^n P(f_i|c) \quad (8)$$

pode agora receber os valores das classes e atributos que nos interessam:

$$\hat{c} = \operatorname{argmax}_{c \in \{\text{spam}, \text{ham}\}} P(c) P(\text{viagra}|c) P(\text{reunião}|c)$$

EXEMPLO DE APLICAÇÃO: CALCULANDO (CONT.)

Por conveniência, retomamos a tabela com os valores já calculados e aplicamos a fórmula que vimos anteriormente para cada classe (*spam* e *ham*) e atributo (palavras).

	spam	ham
N	0,6	0,4
viagra	0,6	0,01
milhões	0,4	0,25
reunião	0,015	0,03

$P(spam)$ 0.6	$P(viagra spam)$ 0.6	$P(reunião spam)$ 0.015	$= 0.0054$
$P(ham)$ 0.4	$P(viagra ham)$ 0.01	$P(reunião ham)$ 0.03	$= 0.00012$

DETALHES DA IMPLEMENTAÇÃO NO MUNDO REAL

- Lembre-se: é esperado que no corpus de testes existam palavras que não estejam no vocabulário de treinamento. Elas devem ser descartadas dos cálculos.

DETALHES DA IMPLEMENTAÇÃO NO MUNDO REAL

- Lembre-se: é esperado que no corpus de testes existam palavras que não estejam no vocabulário de treinamento. Elas devem ser descartadas dos cálculos.
- É importante realizar os cálculos de probabilidades encadeadas usando logaritmos para evitar o *underflow* aritmético. A conversão da equação usada é simples. Repare que, com logartimos, as probabilidades são somadas ao invés de multiplicadas:

$$\hat{c} = \operatorname{argmax}_{c \in \mathcal{C}} \log P(c) + \sum_{i=1}^n \log P(f_i | c) \quad (9)$$

DETALHES DA IMPLEMENTAÇÃO NO MUNDO REAL

- Lembre-se: é esperado que no corpus de testes existam palavras que não estejam no vocabulário de treinamento. Elas devem ser descartadas dos cálculos.
- É importante realizar os cálculos de probabilidades encadeadas usando logaritmos para evitar o *underflow* aritmético. A conversão da equação usada é simples. Repare que, com logaritmos, as probabilidades são somadas ao invés de multiplicadas:

$$\hat{c} = \operatorname{argmax}_{c \in \mathcal{C}} \log P(c) + \sum_{i=1}^n \log P(f_i | c) \quad (9)$$

- Por fim, é também importante empregar a **suavização de Laplace** às probabilidades dos atributos, somando-se 1 à contagem dos atributos no numerador e o tamanho do vocabulário no denominador.

Classificadores

Classificadores bayesianos

Exemplo de aplicação

Avaliação

TIPOS DE AVALIAÇÃO

- Há dois tipos de avaliação de modelos linguísticos:

TIPOS DE AVALIAÇÃO

- Há dois tipos de avaliação de modelos linguísticos:
 - Avaliação Intrínseca

TIPOS DE AVALIAÇÃO

- Há dois tipos de avaliação de modelos linguísticos:
 - Avaliação Intrínseca
 - Avaliação Extrínseca

TIPOS DE AVALIAÇÃO

- Há dois tipos de avaliação de modelos linguísticos:
 - Avaliação Intrínseca
 - Avaliação Extrínseca
- Na avaliação intrínseca, os modelos treinados são confrontados em um **conjunto de testes**.

TIPOS DE AVALIAÇÃO

- Há dois tipos de avaliação de modelos linguísticos:
 - Avaliação Intrínseca
 - Avaliação Extrínseca
- Na avaliação intrínseca, os modelos treinados são confrontados em um **conjunto de testes**.
- A fim de testar o classificador, o corpus é dividido em dois:

TIPOS DE AVALIAÇÃO

- Há dois tipos de avaliação de modelos linguísticos:
 - Avaliação Intrínseca
 - Avaliação Extrínseca
- Na avaliação intrínseca, os modelos treinados são confrontados em um **conjunto de testes**.
- A fim de testar o classificador, o corpus é dividido em dois:
 - Treinamento: 80% dos documentos

TIPOS DE AVALIAÇÃO

- Há dois tipos de avaliação de modelos linguísticos:
 - Avaliação Intrínseca
 - Avaliação Extrínseca
- Na avaliação intrínseca, os modelos treinados são confrontados em um **conjunto de testes**.
- A fim de testar o classificador, o corpus é dividido em dois:
 - Treinamento: 80% dos documentos
 - Teste: 20% dos documentos

TIPOS DE AVALIAÇÃO

- Há dois tipos de avaliação de modelos linguísticos:
 - Avaliação Intrínseca
 - Avaliação Extrínseca
- Na avaliação intrínseca, os modelos treinados são confrontados em um **conjunto de testes**.
- A fim de testar o classificador, o corpus é dividido em dois:
 - Treinamento: 80% dos documentos
 - Teste: 20% dos documentos
- Na avaliação extrínseca, são efetivamente implementados na solução de tarefas. A avaliação é dada por comparação com o desempenho anterior na tarefa usando algum outro modelo ou um baseline.

MATRIZ DE CONFUSÃO

Uma matriz de confusão é uma tabela que exhibe os resultados (outputs) de um algoritmo e permite compará-los aos resultados esperados (dados por etiquetagem, em geral).

		etiquetado	
		<i>spam</i>	<i>ham</i>
estimado	<i>spam</i>	VP	FP
	<i>ham</i>	FN	VN

MÉTRICAS DE AVALIAÇÃO

As métricas de avaliação mais frequentemente usadas são:

MÉTRICAS DE AVALIAÇÃO

As métricas de avaliação mais frequentemente usadas são:

- Acurácia: fração de acertos do classificador

$$\frac{VP + VN}{VP + FP + VN + FN} \quad (10)$$

MÉTRICAS DE AVALIAÇÃO

As métricas de avaliação mais frequentemente usadas são:

- Acurácia: fração de acertos do classificador

$$\frac{VP + VN}{VP + FP + VN + FN} \quad (10)$$

- Precisão: fração das mensagens classificadas como spam que, de fato, eram spam.

Um classificador com alta precisão dificilmente põe um dado em uma classe à qual ele não pertence.

$$\frac{VP}{VP + FP} \quad (11)$$

MÉTRICAS DE AVALIAÇÃO

As métricas de avaliação mais frequentemente usadas são:

- Acurácia: fração de acertos do classificador

$$\frac{VP + VN}{VP + FP + VN + FN} \quad (10)$$

- Precisão: fração das mensagens classificadas como spam que, de fato, eram spam.
Um classificador com alta precisão dificilmente põe um dado em uma classe à qual ele não pertence.

$$\frac{VP}{VP + FP} \quad (11)$$

- Cobertura (*recall*): fração das mensagens que, de fato, eram spam e que foram classificadas como tal. Um classificador com boa cobertura raramente deixa de incluir um dado em uma classe a que ele pertence.

$$\frac{VP}{VP + FN} \quad (12)$$

MÉTRICAS DE AVALIAÇÃO

As métricas de avaliação mais frequentemente usadas são:

- Acurácia: fração de acertos do classificador

$$\frac{VP + VN}{VP + FP + VN + FN} \quad (10)$$

- Precisão: fração das mensagens classificadas como spam que, de fato, eram spam.
Um classificador com alta precisão dificilmente põe um dado em uma classe à qual ele não pertence.

$$\frac{VP}{VP + FP} \quad (11)$$

- Cobertura (*recall*): fração das mensagens que, de fato, eram spam e que foram classificadas como tal. Um classificador com boa cobertura raramente deixa de incluir um dado em uma classe a que ele pertence.

$$\frac{VP}{VP + FN} \quad (12)$$

- Medida-F (*F-score* ou *F-1*): combina *precisão* e *cobertura*.

$$F = 2 \times \frac{\text{precisão} \times \text{cobertura}}{\text{precisão} + \text{cobertura}} \quad (13)$$

MÉTRICAS DE AVALIAÇÃO

As métricas de avaliação mais frequentemente usadas são:

- Acurácia: fração de acertos do classificador

$$\frac{VP + VN}{VP + FP + VN + FN} \quad (10)$$

- Precisão: fração das mensagens classificadas como spam que, de fato, eram spam.
Um classificador com alta precisão dificilmente põe um dado em uma classe à qual ele não pertence.

$$\frac{VP}{VP + FP} \quad (11)$$

- Cobertura (*recall*): fração das mensagens que, de fato, eram spam e que foram classificadas como tal. Um classificador com boa cobertura raramente deixa de incluir um dado em uma classe a que ele pertence.

$$\frac{VP}{VP + FN} \quad (12)$$

- Medida-F (*F-score* ou *F-1*): combina *precisão* e *cobertura*.

$$F = 2 \times \frac{\text{precisão} \times \text{cobertura}}{\text{precisão} + \text{cobertura}} \quad (13)$$

MÉTRICAS DE AVALIAÇÃO

As métricas de avaliação mais frequentemente usadas são:

- Acurácia: fração de acertos do classificador

$$\frac{VP + VN}{VP + FP + VN + FN} \quad (10)$$

- Precisão: fração das mensagens classificadas como spam que, de fato, eram spam.
Um classificador com alta precisão dificilmente põe um dado em uma classe à qual ele não pertence.

$$\frac{VP}{VP + FP} \quad (11)$$

- Cobertura (*recall*): fração das mensagens que, de fato, eram spam e que foram classificadas como tal. Um classificador com boa cobertura raramente deixa de incluir um dado em uma classe a que ele pertence.

$$\frac{VP}{VP + FN} \quad (12)$$

- Medida-F (*F-score* ou *F-1*): combina *precisão* e *cobertura*.

$$F = 2 \times \frac{\text{precisão} \times \text{cobertura}}{\text{precisão} + \text{cobertura}} \quad (13)$$