

Questão 1

Incompleto

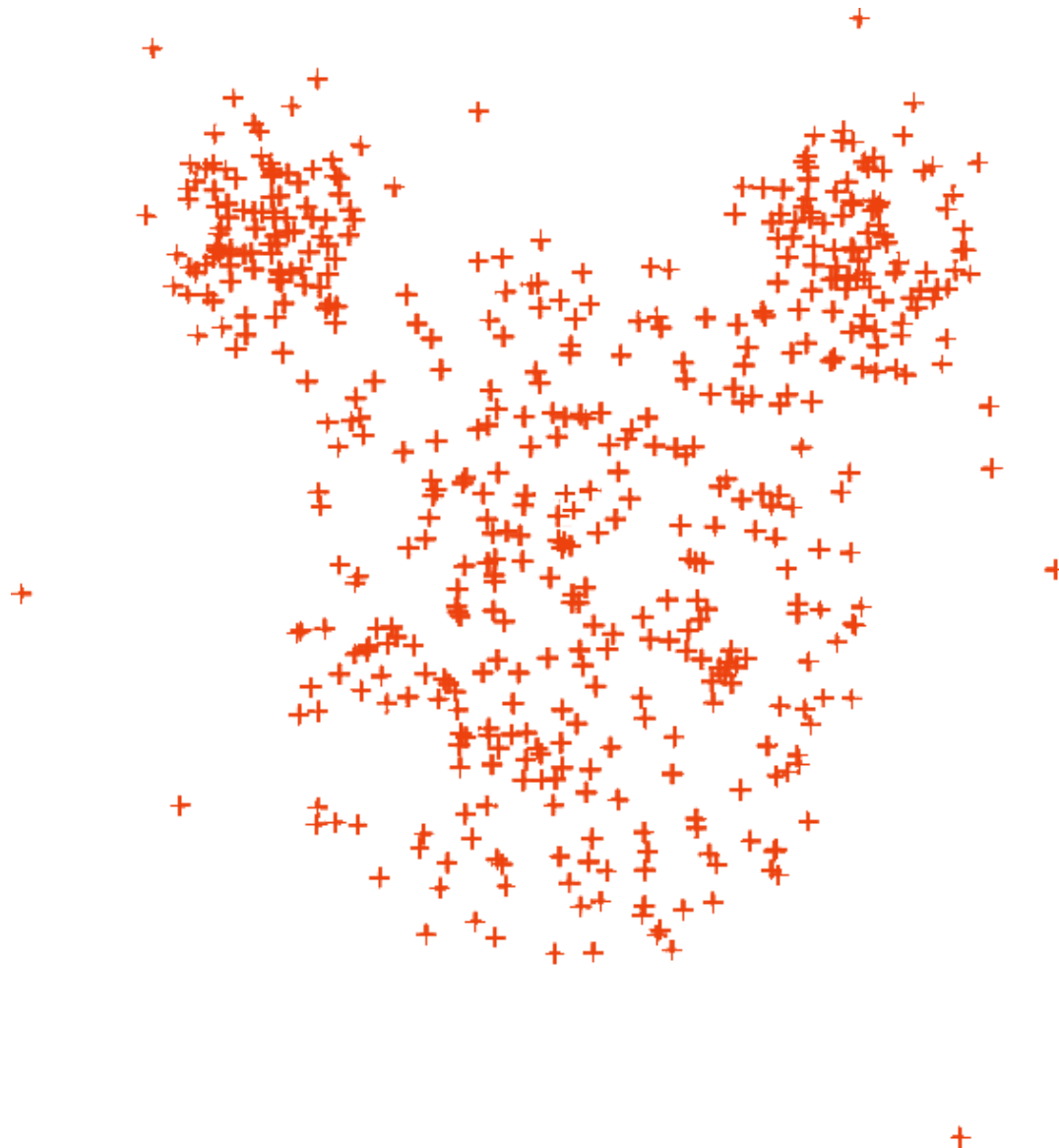
Vale 10,00 ponto(s).

Em busca da amostra mais representativa

Introdução

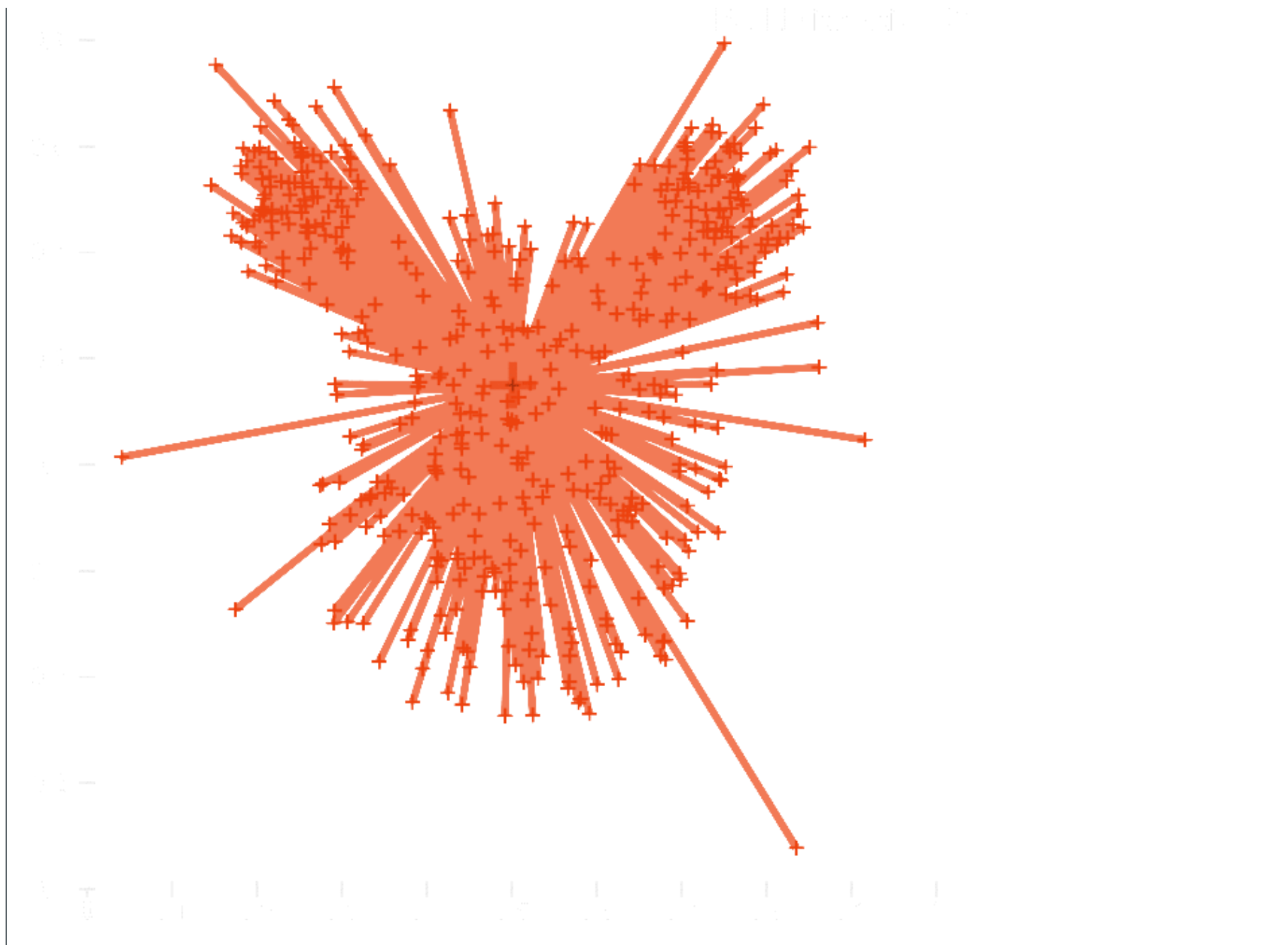
Neste projeto, você irá trabalhar com dados que tem a mesma origem que o projeto anterior: a [Scene Attribute Dataset](#). O objetivo é usar os dados textuais desse conjunto de dados para criar vetores que descrevem os arquivos de imagens, possibilitando o uso de algoritmos de aprendizado de máquina nesses dados. O foco neste projeto é analisar os dados para se obter um vetor que caracterize todo o conjunto.

A ideia é que os objetos presentes em cada imagem sejam usados para a construção de um vetor que a descreve. Na figura a seguir, cada ponto representa um vetor que descreve uma imagem. Chamaremos esse elemento (vetor) de amostra. Essa ilustração é bidimensional, mas na prática usamos vetores com muito mais dimensões. O objetivo final deste trabalho é identificar o vetor mais representativo de todo o conjunto de amostras.



Uma maneira de fazer isso é simplesmente calcular a média (ou centro de massa) de todas as amostras -- que é o que deve ser feito para a tarefa 3. Porém, há algumas desvantagens nisso. Por exemplo, o centro de massa pode ser afetado pela presença de vetores ruidosos (*outliers*), tais como aqueles que ficam bem distantes de todo o resto. Além disso, trata-se de um vetor sintético, construído simplesmente pela média das variáveis. Para algumas aplicações, é bem melhor interpretar o resultado com base na seleção da

amostra real que seja a mais representativa. Isso é possível calculando a distância entre todos os pares de amostras e a identificação de qual delas tem, na média, a menor distância a todas outras, como ilustrado na figura abaixo -- isso é o que deve ser feito para a tarefa 4.



Informações relevantes

- **Propagação de erros:** Assim como no projeto anterior, aconselhamos que sempre que for executada uma sequência de cálculos, execute-a de maneira a minimizar a propagação de erros. Por exemplo, para se calcular a média entre as variáveis A, B e C, calcule desta forma: $(A+B+C)/3$, e não desta: $A/3 + B/3 + C/3$. Apesar de serem matematicamente equivalentes, na segunda forma há mais propagação de erros.
- **Tipo de dados:** Apesar dos dados serem fornecidos como números inteiros, é importante que todas operações matemáticas sejam feitas com números de ponto flutuante. Ao imprimir dados numéricos como valores sem ponto flutuante, não converta dados para inteiro. Use f-strings e imprima seus números como sendo números de ponto flutuante, mas sem casas decimais.
- É permitido o uso de [bibliotecas padrão do Python](#). Note que numpy, pandas e scikit-learn **não** estão incluídas entre as bibliotecas padrão.

Entrada

Os dados são fornecidos exatamente no mesmo formato que no [projeto 1](#):

```
T N
nome_do_arquivo_1
atributo-objeto_1
x1 y1 x2 y2

nome_do_arquivo_2
atributo-objeto_2
x1 y1 x2 y2

.
.
.
nome_do_arquivo_N
atributo-objeto_N
x1 y1 x2 y2
```

Tarefas e saída de dados

As tarefas 1 e 2 são tarefas relacionadas à construção de vetores que descrevem as amostras (imagens). A tarefa 3 se refere a computar a média (centro de massa) desses vetores e a tarefa 4 se refere a identificar qual das amostras é a mais representativa de todas elas.

1. Vetor de indicadores de objetos (T=1)

Reformatar os dados de forma que eles fiquem mais úteis para o método a ser implementado. Para cada imagem, imprima apenas uma linha contendo o nome do arquivo seguido por 16 números binários (separados por um espaço em branco). Cada número indica a presença/ausência de objetos (independentemente da cor ou ponto de vista do objeto), nesta ordem:

1. bison
2. elephant
3. horse
4. ibis
5. sky
6. mountain
7. building
8. flower
9. sand
10. tree
11. field
12. road
13. tower
14. ocean
15. cliff
16. waterfall

Saída

Para esta tarefa, a saída de dados tem o seguinte formato:

```
nome_do_arquivo1 x1 x2 x3 ... x16
nome_do_arquivo2 x1 x2 x3 ... x16
.
.
.
nome_do_arquivoF x1 x2 x3 ... x16
```

Onde x_1 , ..., x_{16} assumem valores 0 ou 1, indicando a ausência ou presença, respectivamente, de cada objeto listado acima na imagem referente ao `nome_do_arquivo`. Note que o número de linhas na saída $F \leq N$, pois em geral cada imagem possui mais de uma caixa de objeto.

2. Vetor de estatísticas das caixas (T=2)

No mesmo formato da tarefa anterior, para cada imagem, imprima estatísticas das caixas (bounding boxes) dos objetos:

1. Total de caixas nesta imagem dividido por 2;
2. Média das coordenadas x dos centros das caixas desta imagem dividida por 128;
3. Média das coordenadas y dos centros das caixas desta imagem dividida por 128;
4. Média das larguras das caixas desta imagem dividida por 128;

5. Média das alturas das caixas desta imagem dividida por 128;
6. Média das áreas das caixas desta imagem dividida por 128^2 (128 ao quadrado);
7. Desvio padrão das coordenadas x dos centros das caixas desta imagem, dividido por 32;
8. Desvio padrão das coordenadas y dos centros das caixas desta imagem, dividido por 32;
9. Desvio padrão das larguras das caixas desta imagem, dividido por 32;
10. Desvio padrão das alturas das caixas desta imagem, dividido por 32.

Importante: assumo que o desvio padrão é a raiz quadrada da variância da população (e não da amostra).

Saída

O formato de saída desta tarefa é bem parecido com o da anterior, exceto que há 10 colunas de números (s_1, \dots, s_{10}) e eles são de ponto flutuante com 1 casa decimal.

```
nome_do_arquivo1 s1 s2 s3 ... s10
nome_do_arquivo2 s1 s2 s3 ... s10
.
.
.
nome_do_arquivoF s1 s2 s3 ... s10
```

3. Vetor médio (T=3)

Use a concatenação dos vetores das duas tarefas anteriores para descrever cada amostra, i.e., para cada imagem, construa um vetor que possui os 16 números da tarefa 1 seguidos pelos 10 números da tarefa 2. Estes vetores serão utilizados nesta e na próxima tarefa.

Calcule e imprima o vetor médio de todas as imagens.

Saída

A saída é uma linha contendo 26 números com 1 casa decimal.

```
m1 m2 m3 ... m26
```

4. Amostra prototípica/*medoid* (T=4)

Para esta tarefa, cada amostra (imagem) é representada pela concatenação do vetor indicadores de presença/ausência de objetos com o vetor de informações estatísticas das caixas (i.e., a combinação das tarefas 1 e 2). Identifique o *medoid* (protótipo) do conjunto, i.e., a amostra cuja [distância L1](#) à todas as demais amostras é a menor possível.

Para o cálculo da distância entre amostras utilize a fórmula abaixo, na qual \mathbf{p} e \mathbf{q} são dois vetores representando as amostras.

$$d_1(\mathbf{p}, \mathbf{q}) = \|\mathbf{p} - \mathbf{q}\|_1 = \sum_{i=1}^n |p_i - q_i|,$$

ou seja, a distância L1 é a soma das diferenças absolutas entre todos os elementos dos dois vetores.

Após o cálculo das distâncias para todos os pares de amostras, calcule para cada amostra a soma das distâncias dela à todas as demais amostras. A amostra que possuir a menor soma é a *medoid* do conjunto.

Saída

A saída é apenas um string, identificando o nome do arquivo da amostra identificada como *medoid*:

nome_do_arquivo

Testes

Uma amostra de casos de teste com o mesmo perfil dos casos usados para avaliar seu trabalho está disponível [neste arquivo](#).

For example:

Input	Result
1 2 horse_073 blue-sky 96 5 242 93 horse_073 brown-horse 36 46 238 195	horse_073 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0

<https://aprender3.unb.br/mod/quiz/attempt.php?attempt=603278&cmid=186271>

Input	Result
highway_gre493 road 17 156 256 256	
opencountry_093 cloudy-sky 1 5 250 99	
opencountry_093 tree 11 99 248 158	
opencountry_093 yellow-field 9 157 255 252	
opencountry_test_030 blue-sky 3 7 249 85	
opencountry_test_030 sand 11 124 254 249	
opencountry_test_030 rocky-mountain 4 69 234 110	
highway_bost148 blue-sky 38 8 241 137	
highway_bost148 road 11 149 251 250	
opencountry_232 red-flower 13 161 251 251	
opencountry_232 green-field 20 116 256 158	

Input	Result
opencountry_232 blue-sky 28 4 239 101	
waterfall_051 green-cliff 115 4 248 236	
waterfall_051 waterfall 64 8 129 256	
waterfall_051 green-cliff 1 4 60 248	
street_par95 dusthaze-sky 104 1 163 97	
street_par95 perspective-building 171 10 256 154	
street_par95 perspective-building 1 2 112 163	
street_par95 road 24 150 256 256	
mountain_068 cloudy-sky 8 5 247 52	
mountain_068 snowy-mountain 24 68 248 222	
whitesand09 overcast-sky 3 9 249 120	

Input	Result
whitesand09 snowy-field 6 132 243 250 tower_033 cloudy-sky 88 12 254 218 tower_033 tower 25 6 79 233 tower_033 green-field 14 234 256 256	
2 3 waterfall_053 waterfall 61 0 159 256 waterfall_053 green-cliff 158 4 254 246 waterfall_053 green-cliff 8 7 80 235	waterfall_053 1.5 0.9 1.0 0.7 1.9 1.3 2.1 0.1 0.4 0.4

14/26

Input	Result
waterfall_070 green-cliff 113 13 251 251	elephant17 2.5 1.2 1.0 1.3 0.7 0.8 1.4 2.2 2.0 1.7 street_par42 1.5 1.1 1.0 1.1 1.1 1.0 1.8 1.8 1.5 1.1 highway_bost291 2.0 1.1 1.0 1.2 0.6 0.8 1.8 1.6 2.4 1.1 opencountry_108 1.0 1.0 1.0 1.8 0.9 1.6 0.0 2.0 0.0 0.0
waterfall_070 green-cliff 5 9 54 239	
opencountry_test_022 cloudy-sky 1 7 250 105	
opencountry_test_022 yellow-field 16 129 248 246	
ocean08 blue-sky 13 7 250 104	
ocean08 blue-ocean 6 134 249 251	
ibis_137 white-ibis 60 10 235 215	
waterfall_063 green-cliff 134 16 255 251	
waterfall_063 waterfall 71 8 126 241	
waterfall_063 green-cliff 3 5 62 247	
city01 birdseye-building 12 71 251 253	

Input	Result
city01 blue-sky 3 1 251 63	
elephant17 elephant 9 45 178 207	
elephant17 green-field 25 187 250 254	
elephant17 rocky-mountain 149 154 252 187	
elephant17 overcast-sky 9 3 249 41	
elephant17 overcast-sky 170 5 252 157	
street_par42 perspective-building 160 9 250 182	
street_par42 perspective-building 9 3 120 156	
street_par42 road 33 164 237 256	
highway_bost291 cloudy-sky 5 0 231 117	
highway_bost291 tree 10 124 140 147	

Input	Result
highway_bost291 road 14 160 246 248 highway_bost291 tree 211 95 253 153 opencountry_108 dusthaze-sky 11 6 246 118 opencountry_108 yellow-field 13 137 247 248	
3 2 waterfall136 waterfall 103 8 252 249 waterfall136 black-cliff 9 14 90 249	0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0 1.0 1.0 0.9 1.0 0.9 1.9 1.7 2.0 0.0 1.1 0.1

Input	Result
3 8 waterfall_070 waterfall 58 0 115 256 waterfall_070 green-cliff 113 13 251 251 waterfall_070 green-cliff 5 9 54 239 volcano_0121 blue-sky 18 10 256 85 volcano_0121 blue-sky 156 10 255 148 volcano_0121 rocky-mountain 5 87 248 252 volcano_034 blue-sky 9 7 256 139 volcano_034 rocky-mountain 5 130 255 253	0.0 0.0 0.0 0.0 0.7 0.7 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.3 0.3 1.3 1.0 0.9 1.4 1.3 1.5 1.0 1.2 1.1 0.6
3 28 street_par29 tree 13 18 115 174 street_par29 tree 138 108 201 180	0.0 0.0 0.2 0.0 0.7 0.1 0.1 0.1 0.0 0.4 0.4 0.4 0.0 0.0 0.1 0.1 1.4 1.1 1.0 1.5 0.9 1.4 1.0 1.8 1.3 0.9

Input	Result
street_par29 perspective-building 195 68 256 195	
street_par29 road 16 156 240 255	
opencountry_115 cloudy-sky 3 2 245 132	
opencountry_115 yellow-field 0 123 251 253	
highway_bost150 blue-sky 8 7 256 148	
highway_bost150 road 7 177 253 250	
highway_bost150 tree 169 146 256 184	
opencountry_001 red-flower 6 125 204 250	
opencountry_001 yellow-field 177 123 254 228	
opencountry_001 blue-sky 2 2 247 121	
opencountry_test_026 blue-sky 1 5 248 94	

Input	Result
opencountry_test_026 yellow-field 4 97 249 254	
horse_055 green-field 7 191 250 250	
horse_055 white-horse 45 58 224 179	
horse_055 tree 0 9 92 50	
horse_055 dusthaze-sky 87 4 253 48	
waterfall119 waterfall 98 17 249 252	
waterfall119 black-cliff 2 6 86 247	
horse_138 tree 97 6 247 64	
horse_138 white-horse 24 29 247 242	
horse_138 road 6 221 243 253	
volcano_0271 dusthaze-sky 12 3 248 94	

Input	Result
volcano_0271 dusthaze-sky 164 6 252 145 volcano_0271 rocky-mountain 6 102 240 252 highway_art1674 blue-sky 9 4 248 134 highway_art1674 road 19 156 250 251	
4 2 opencountry_072 sand 10 85 253 251 opencountry_072 cloudy-sky 12 8 250 64	opencountry_072

Input	Result
4 9 highway_bost300 blue-sky 4 5 249 124 highway_bost300 tree 5 114 127 177 highway_bost300 road 25 186 254 250 ibis_072 green-field 19 214 243 253 ibis_072 black-ibis 56 41 239 179 waterfall36 waterfall 103 8 252 249 waterfall36 black-cliff 9 14 90 249 opencountry_023 snowy-field 5 123 251 252 opencountry_023 blue-sky 10 8 247 117	opencountry_023
4 29 opencountry_159 blue-sky 4 1 252 60	bison_045

Input	Result
opencountry_159 rocky-mountain 4 69 256 102	
opencountry_159 sand 10 117 253 253	
tower_066 tree 6 3 140 192	
tower_066 tower 149 7 242 205	
tower_066 green-field 24 195 256 255	
tower27 tower 221 20 251 242	
tower27 overcast-sky 3 14 211 190	
tower27 green-field 6 208 207 253	
horse_049 brown-horse 52 54 236 217	
horse_049 green-field 13 4 253 79	
horse_049 green-field 8 129 90 251	

Input	Result
horse_049 green-field 8 174 256 254	
tower02 overcast-sky 17 5 188 197	
tower02 tower 202 7 247 197	
tower02 blue-ocean 6 208 253 252	
waterfall123 black-cliff 4 8 95 248	
waterfall123 waterfall 94 8 255 244	
bison_045 black-bison 64 72 248 207	
bison_045 blue-sky 2 1 253 58	
bison_045 green-field 0 203 253 255	
opencountry_test_031 overcast-sky 7 1 253 109	
opencountry_test_031 sand 9 127 252 251	

Input	Result
opencountry_239 red-flower 9 96 256 256	
opencountry_239 tree 57 53 244 98	
opencountry_239 overcast-sky 0 5 250 57	
opencountry_048 overcast-sky 6 6 254 82	
opencountry_048 rocky-mountain 23 98 249 152	
opencountry_048 green-field 11 169 245 245	

Answer: (penalty regime: 0, 0, 10, 20, ... %)

1 ||

Precheck

Verificar

<<